Motif-*h*: a novel functional backbone extraction for directed networks

Yiguang Bai¹ · Qian Li¹ · Yanni Fan¹ · Sanyang Liu¹

Received: 3 October 2020 / Accepted: 7 August 2021 / Published online: 18 September 2021
© The Author(s) 2021

Abstract

Dense networks are very pervasive in social analytics, biometrics, communication, architecture, etc. Analyzing and visualizing such large-scale networks are significant challenges, which are generally met by reducing the redundancy on the level of nodes or edges. Motifs, patterns of the higher order organization compared with nodes and edges, are recently found to be the novel fundamental unit structures of complex networks. In this work, we proposed a novel motif *h*-backbone (Motif-*h*) method to extract functional cores of directed networks based on both motif strength and *h*-bridge. Compared with the state-of-the-art method Motif-DF and Entropy, our method solves two main issues which are often found in existing methods: the Motif-*h* reconsiders weak ties into our candidate set, and those weak ties often have critical functions of bridges in networks; moreover, our method provides a trade-off between the motif size and the edge strength, which quantifies the core edges accordingly. In the simulations, we compare our method with Motif-DF in four real-world networks and found that Motif-*h* can streamline the extraction of crucial structures compared with the others with limited edges.

Keywords Backbone extraction · Salient motifs · Motif *h*-strength · *h*-Bridge

Introduction

Modern network science has progressed rapidly and has become an important research tool in sociology, computer science, biology, etc. [4,7,9,13,28,44]. The key to understanding the network is in-depth mining its intrinsic topology and finding the most represent backbone from a great number of complex connections. As the real-world networks tends to be highly dense and large scale, how to extract the fundamental features and reduce redundant of dense networks efficiently become one of the most popular issues in network science [19,39,42,55].

Extracting the so-called backbone of a network is a very challenging problem, which relies on the concept of minimizing large-scale networks while still preserves the essential

information of the original network [14]. Intuitively, nodes with a higher centrality value will be prior preserved as the core structure. With this respect, numerous researches and excellent works have proposed to handle such problem [46,51], which can be classified into three categories: coarse-graining model, filtering model and generalist model.

Coarse-graining methods reduce the network size by grouping similar nodes, and they consider these groups as a single node [18]. In this sense, communities can be replaced by a single node under a coarse view [16,53]. Song et al. propose the box-counting techniques to reduce the sizes of networks [43]. Zeng et al. proposed the new path-based algorithm to coarse-grain the directed networks, which can effectively preserve the synchronizability of the directed networks [52]. The most typical work is the spectral coarse-graining (SCG) method, proposed by Gfeller et al., which is mainly based on the relationships between the eigenvalue spectrum of the Laplacian matrix of the network and the synchronizability of the network to merge similar nodes [16,17]. One significant weakness of these methods is that the extraction results highly depend on the community defined, where proper communities findings is an NP-hard problem [6].

Filter-based methods prune the network by discarding nodes or edges based on a statistical property. Generalist model can be viewed as the special term of filter-based

✉ Qian Li
qianli_30@163.com

Yiguang Bai
ygbai@foxmail.com

Yanni Fan
fanyanni36@foxmail.com

Sanyang Liu
liusanyang@126.com

¹ School of Mathematics and Statistics, Xidian University, Xi'an, Shaanxi, China

method but focus on the edge's dimensionality cutting. Previous researches achieve the backbone's extractions by applying a hard threshold according to the node attributes, such as the degree, betweenness, coreness centrality, etc. [6,25,29,41]. Grady et al. have introduced the robust link salience method to extract network skeletons of generic statistical properties based on the shortest path tree [20]. Yuan et al. have proposed the TP_{ks} index to identify the important nodes in the network and achieved the success in the analysis of the bicycle sharing networks [30,37]. Zhang and Zhu have proposed the new measure, i.e., the strong ties, as skeletons of weighted social networks [56]. While researchers have also found that social network structures are robust to removing strong relationships, yet gradually falling apart as removing the weak ties [21]. Besides the above methods, disparity filter (DF), locally adaptive network sparsification (LANS), and globally and locally adaptive network backbone (GLANB), commonly extract the significant edges, according to the null model-based weights in networks [39,42,55].

All above methods follows one same rule: nodes or edges with scores higher than the predetermined values are preserved as the final core structure. However, these measures are limited in their performances to describe non-linear, localized, and dynamic properties of the systems from the views of nodes and edges. To uncover structural, temporal, and functional insights of complex systems, network motifs have been used extensively in recent years as they provide a tractable approximation of the networks that can be measured and updated within given memory and compute constraints [38]. Network motifs are patterns of interactions occurring in the complex system at a rate higher than those in a randomized network. To the best of our knowledge, most backbone extraction methods have been obtained only at the level of nodes or edges. In contrast, few attempts have been made here to abstract functional backbone except for Cao [12]. New measure motifs are the fully represented subgraph with higher-order connectivity patterns in networks [11], which are the basic building blocks that manipulate the activity of most complex systems [3]. Cao et al. first proposed a motif-based disparity filter, i.e., Motif-DF, in which he uses normalized motif weight of edges to extract functional backbone. However, when most of the edge weights connected to a certain vertex are large, partially normalized weights in Motif-DF will lower the significant global advantage of these edges and miss very important skeletons in the final result. Moreover, Motif-DF highly relies on the external static parameter named significance level α , to determine the number of edges preserved and performance in the backbone detection.

Inspired by the recent works [12,54,58], we introduce a novel motif h -backbone (Motif- h) method for directed networks, which combines and h -index method and weak ties theory into the significant motif edges finding. The new

method will be designed to establish an adaptive h -index for each network, free from the constraint of fixed parameters, e.g. α in Motif-DF. Besides, the new global normalization strategy preserves the advantages of edges and facilitates the optimal global selection of edges. Considering weak ties, we adopt the h -bridge in the second selection stage, where the weak ties have been ignored by the existing method, i.e., Motif-DF. The newly proposed Motif- h method will provide a particular perspective to study the distribution of motif edge strength and networks' functional structure. At the same time, the Motif- h method simplifies the networks and preserves the vital functional structures effectively.

The outline of this work is as follows: the next section introduces the definitions related to the motifs and the motif backbone problem. In the subsequent section, the motivation and details of our Motif- h method are presented; then the performance among three methods, i.e., Motif- h , Motif-DF, and Entropy in various real-world networks is presented. The last section summarizes our work and shows the future directions.

Problem formulation and methodology

Motifs are the fundamental functional structures which can give a better perspective to make it easy to understand and analyze large-scale complex networks. One breakthrough work shows [35] that network motifs can build the blocks of complex networks, and different types of networks have distinct in the motif preserving [34,36]. That means the combinations of motifs represent certain essential features in the network compared to nodes and edges.

Comparing with traditional backbone methods with the nodes or edges, the functional backbones with the related motifs preserved will extend the ability to understand structural, nonlinear information inside the complex topology. In this section, we will introduce the functional backbone problem based on the motifs, and the key notations are listed in Table 1.

Salient motifs and functional backbone problem

To illustrate the characters of motifs, we list 13 three-node motifs in Fig. 1, which have been widely found in various networks. Typically, different motifs serve distinct functions in different networks [33]. For example, the triangular motifs ($M_1 - M_7$) that are statistically significant exist in social networks; the feedforward loops M_5 are fundamental to transcriptional regulation networks and neural networks [31,57]; the two-hop paths ($M_8 - M_{13}$) help us to understand the travel patterns of air traffics [22,40]; the open bidirectional wedges (M_{13}) are the most potent motif for the information flow in functional brain networks [32]. M_5 and M_9 are tri-

Table 1 Notations and definitions

Symbol	Description
$\mathcal{A}$	The adjacency matrix
$\mathcal{G}$	The network
$M_k, k = 1, 2, \dots, 13$	The three-node Motifs in networks
$\mathbb{M}$	The salient motif set
o_{ij}^k	The motif edge occurrences according to M_k
w_{ij}	The motif edge strength strength
$\hat{w}_{ij}$	The normalized edge strength
$h_{\mathbb{M}}$	The h -index
$\mathcal{B}(e_{ij})$	The normalized bridge strength
MD	The total motif degree in the network
MC($\mathbb{M}$)	The total motif centrality in the network

Fig. 1 All 13 three-node connected motifs of directed networks [35]

trophic food chains and omnivory chains, respectively, which represent the relationships among predators, consumers and resources in ecological networks [8,48].

One famous open-source software, i.e., mfindert [34], gives a way to detect the motifs in various types of networks. For directed networks, there are 13 three-node motifs and 199 four-node motifs [45]. To simplify the calculations, we focus our research area on the motifs with three nodes. Basically, not every motif is of the same significance in real-world networks. Salient motifs are those whose occurrence is no less than that of the ensemble of randomized networks with identical degree sequences [34]. Therefore, we use the significance profile ($\mathcal{SP}$) approach to analyze the motifs with a much higher occurrence frequency than the expected [34].

Given a directed network $\mathcal{G} := (\mathcal{V}, \mathcal{E}, \mathcal{N})$, with $\mathcal{N}$ vertices (or nodes) and $|\mathcal{E}| = L$ edges (or links). Denote the significance of a motif M_k as Z score:

$$\mathcal{Z}(k) = \frac{N_k^{\text{real}} - \langle N_k^{\text{rand}} \rangle}{\sigma(N_i^{\text{rand}})}, \quad (1)$$

where N_k^{real} is the occurrences of the motif M_k in real networks, $\langle N_k^{\text{rand}} \rangle$ and $\sigma(N_i^{\text{rand}})$ are the mean and standard deviation of the appearances of the motif M_k in its corresponding randomized network ensemble. Then, the $\mathcal{SP}$ value of M_k is defined as the normalized Z score:

$$\mathcal{SP}(k) = \frac{\mathcal{Z}(k)}{\sum_k \mathcal{Z}(k)} \quad k = 1, 2, \dots, 13. \quad (2)$$

As we want to fully explore the complex network $\mathcal{G}$, it is necessary to carefully construct the network structures and functions using salient motifs. To determine every salient motif M_k from the according motif, we generate a motif edge strength matrix (see as Definition 2) to measure the functional strength of each edge [11]. Clearly, higher motif edge strength means that some pairs of nodes have stronger functional interactions than others. Thus, the salient motif set is denoted as

$$\mathbb{M} = \{M_k | \mathcal{SP}(k) > 0, k = 1, \dots, 13\}, \quad (3)$$

in which motifs with the $\mathcal{SP}$ values greater than 0 are referred to as salient motifs.

Definition 1 *Motif edge occurrences:* Given a specific salient motif $M_k \in \mathbb{M}$, there is a related $M_k = \{M_k^1, M_k^2, \dots, M_k^l\}$ set include all motif M_k in the current network, the motif edge occurrences according to M_k is defined as

$$o_{ij}^k = \sum_{c=1}^l P_{kc}(e_{ij}); \quad (4)$$

where

$$P_{kc}(e_{ij}) = \begin{cases} 1, & e_{ij} \in M_k^c \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

Definition 2 *Motif edge strength matrix:* Given a specific salient motif M_k , the edge strength matrix $W = (w_{ij})$ is defined as the total occurrence count of edge e_{ij} :

$$w_{ij} = \sum_{M_k \in \mathbb{M}} o_{ij}^k = \sum_{M_k \in \mathbb{M}} \sum_{M_k^c \in M_k}^l P_{kc}(e_{ij}). \quad (6)$$

In this paper, we aim at solving the problem of extracting the functional backbone from a dense complex network. A functional backbone is informally defined as a subset of nodes and edges of the original network which contains the most significant motifs with certain nodes and edges reserved.

Problem Definition: Given a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{N})$, find a smallest sub-graph $\mathcal{G}^* = (\mathcal{V}^*, \mathcal{E}^*, \mathcal{N}^*)$ maintains the highest motif degree (Definition 5) or motif centrality (Definition 6), such that $\mathcal{E}^* \subset \mathcal{E}$, $\mathcal{V}^* \subset \mathcal{V}$ and follows:

- (a) $|\mathcal{MN}_{\mathcal{G}^*}| \leq |\mathcal{MN}_{\mathcal{G}}|$, where $\mathcal{MN}_{\mathcal{G}}$ is the number of Motifs. (non-increase)
- (b) $\mathcal{SP}_{\mathcal{G}^*}(k) \geq \mathcal{SP}_{\mathcal{G}}(k)$, where $M_k \in \mathbb{M}$. (strengths enhance)
- (c) $\mathcal{Z}_{\mathcal{G}^*}(k) \rightarrow -\frac{\langle N_k^{\text{rand}} \rangle}{\sigma(N_i^{\text{rand}})}$, where $M_k \notin \mathbb{M}$. (noisy disparity)

The motif-based disparity filter (Motif-DF) method

To solve the motif-based backbone extraction problem, Cao [12] gives the first attempt and proposed the motif based disparity filter (Motif-DF) method, where the edges should be preserved as backbone by comparing their weights via the null model, and the probability of the edges' normalized weight as follows:

$$\tau_{ij} = 1 - \int_0^{\hat{w}_{ij}} (k-1)(1-x)^{k-2} dx, \quad (7)$$

where k is the degree of the source node, x is the particular value, and here we use the normalized motifs weight instead. The normalized motifs weight $\hat{w}_{ij}$ is defined as

$$\hat{w}_{ij} = \frac{w_{ij} + a_{ij}}{\sum_j (w_{ij} + a_{ij})}, \quad (8)$$

where a_{ij} is the element of adjacency matrix $\mathcal{A}$, and the Motif-DF method preserves the nodes when $\tau_{ij} \leq \alpha$.

Algorithm 1 Motif-DF Method

Input: Adjacency matrix $\mathcal{A}$ of the network $\mathcal{G}$, the edge set $\mathcal{E}$.

Output: the selected Motif h -backbone set Ω .

```

1: Set  $\Omega = \emptyset$ , and calculate the motif edge strength matrix  $W = (w_{ij})$ 
   by Eq. 6.
2: Filter the salient motifs set by Eq. 1.
3: for each salient motif  $M_k \in \mathbb{M}$  do
4:   Calculate the normalized motif weight  $\hat{w}_{ij}$  by Eq. 8.
5:   for each edge  $e_{ij}$  in the network  $\mathcal{G}$  do
6:     if  $\tau_{ij} \leq \alpha$  then
7:       Add the edge  $e_{ij}$  to  $\Omega$ .
8:     end if
9:   end for
10: end for
11: return  $\Omega$ 
```

Motif- h method

h -strength and h -bridge

To solve the functional backbone extraction problem, we introduce a novel motif- h method which derives from the filter-based method but fully focus on silent motifs by h -index methods, i.e., h -strength and h -bridge. Considering the partial normalization in Motif-DF fails in selecting edges (see Eq. 8), as the strength weights of most edges connect to specific nodes are very high. This paper constructs the new h -strength with the novel global normalization, which will retain the edges' original advantage and map the weights domain ($[0 \ 1]$) to the same domain as the nodes' number ($[0 \ \mathcal{N}]$):

$$\tilde{w}_{ij} = \frac{\mathcal{N}w_{ij}}{\max(w_{ij})}. \quad (9)$$

With respect to the motifs in networks, we first define a metric to measure the backbone nodes in terms of edges' strength and quantities, a.k.a motif h -strength:

Definition 3 *Motif h -strength:* Given a network $\mathcal{G}$, the motif h -strength is equal to $h_{\mathbb{M}}$, if $h_{\mathbb{M}}$ is the largest natural number such that there are $h_{\mathbb{M}}$ edges each with motif edge strength larger than or equal to $h_{\mathbb{M}}$. For the fully descent strength order edges set $\{e_1, e_2, \dots, e_S\}$ in all salient motifs $\mathbb{M}$, we have its h -strength index form as

$$h_{\mathbb{M}} = \max \{h \mid h \geq s, h \geq \tilde{w}_{e_s}, h \in \mathbb{N}, s \in 1, 2, \dots, S\}. \quad (10)$$

To make it more adaptive to various requirements, we extend the h -strength, as the fractional- h form:

$$h_{\mathbb{M}}^\beta = \max \{h \mid h \geq s, h \geq \tilde{w}_{e_s}^\beta, h \in \mathbb{N}, s \in 1, 2, \dots, S\}. \quad (11)$$

Most of the current methods require an additional parameter to guide the number of filters, and the setting of the parameter is often not related to the nature of the network itself. Different networks have differentiated network properties due to complex connectivity relationships, and obviously, it is not wise to pick the same parameter to guide. With this respect, we newly proposed the motif h -strength, which provides a particular measure to analyze edges with high strength and edge strength distributions corresponding to each salient motif. In the Motif- h method, we only need to calculate the h -index of each network corresponding to the preservation of all edges larger than $h_{\mathbb{M}}$, and no longer need an additional parameter intervention.

h -strength indicates that motifs with higher occurrences are more likely to have higher motif edge strength than others. Actually, besides the high weighted edges, weak ties also play a vital role as bridges in a network and the edge betweenness measures the total amount of an edge functioning as a bridge in a network. Naturally, edges with higher edge betweenness value are more critical. However, motif edge strength cannot find out those edges with low motif-strength but high betweenness value. In keeping with the motif h -strength, here we use the h -bridge to measure those critical edges [54]. In the network $\mathcal{G}$, the bridge strength of an edge e_{ij} is defined as

$$\mathcal{B}(e_{ij}) = \sum_{v_s \neq v_t \in V} \frac{\delta_{st}(e_{ij})}{\mathcal{N} \delta_{st}} = \frac{\text{BCL}(e_{ij})}{\mathcal{N}}, \quad (12)$$

where $\mathcal{N}$ is the size of the network, δ_{st} is the number of all shortest paths from node s to node t , $\delta_{st}(e_{ij})$ is the total number of shortest paths passing through the edge e_{ij} . $\text{BCL}(e_{ij})$ is the edge betweenness, which is defined as the total number of shortest paths that pass through e_{ij} [5] from all vertices to all the other vertices, such that

$$\text{BCL}(e_{ij}) = \sum_{v_s \neq v_t \in V} \frac{\delta_{st}(e_{ij})}{\delta_{st}}. \quad (13)$$

Since $\text{BCL}(e_{ij})$ value ranges in $[0 \mathcal{N}^2]$ in a network. In this work, we use Eq. (12), so as to map the range to $[0 \mathcal{N}]$. Besides the Motif h -strength, we also adopt the h -index approach to rank the bridges in the network, then we get the h -bridge [54], such that

Definition 4 h -bridge: The h -bridge of a network is equal to h_b , if h_b is the largest natural number such that there are h_b edges each with a bridge at least equal to h_b in the given network. With the fully descent strength order edges set $\{e_1, e_2, \dots, e_P\}$ in current network, and its related bridge strength $\{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_P\}$, we have its h -bridge index form as

$$h_b = \max \{h \mid h \geq p, h \geq \mathcal{B}_p, h \in \mathbb{N}, p \in 1, 2, \dots, P\}. \quad (14)$$

Here, we also give the Fractional- h form for the bridge edges:

$$h_b^\beta = \max \{h \mid h \geq p, h \geq \mathcal{B}_p^\beta, h \in \mathbb{N}, p \in 1, 2, \dots, P\}. \quad (15)$$

Apparently, the motif h -strength quantifies the functionally significant edges for different salient motifs. Relatively, h -bridge characterizes the structurally critical edges whose removal disconnects the network.

Considering the weak ties and difference among the motifs, in this paper, we propose the Motif- h method to extract the backbones from the directed network, which is composed of all edges with motif edge strength larger than or equal to the $h_{\mathbb{M}}$ and with bridges larger than or equal to h_b , and the nodes connected by these edges. The details for the algorithm of Motif- h are shown in Algorithm 2, which consists of two parts: strong motif-weight edge extraction and weak ties (strong bridging) edge extraction:

Algorithm 2 Motif h -Backbone (Motif- h) Method

Input: Adjacency matrix $\mathcal{A}$ of the network $\mathcal{G}$, the edge set $\mathcal{E}$.

Output: the selected Motif h -backbone set Ω .

1: Set $\tilde{\Omega} = \emptyset$, and calculate the motif edge strength matrix W .

2: Filter the salient motifs set by Eq. 1.

3: **for** each salient motif $M_k \in \mathbb{M}$ **do**

4: Calculate the motif h -strength $h_{\mathbb{M}}$ by Eq. 11.

5: **for** each edge e_{ij} in the network $\mathcal{G}$ **do**

6: **if** $w_{ij}^k \geq h_{\mathbb{M}}$ **then**

7: Add the edge e_{ij} to $\tilde{\Omega}$.

8: **end if**

9: **end for**

10: **end for**

11: **for** each edge e_{ij} in the network $\mathcal{G}$ **do**

12: Calculate the bridge $\mathcal{B}(e_{ij})$.

13: **end for**

14: Calculate the h -bridge h_b by Eq. 15.

15: Set $\Omega_b = \emptyset$ where Ω_b is the subset of Ω with bridge function

16: **for** each edge e_{ij} in the network $\mathcal{G}$ **do**

17: **if** $\mathcal{B}(e_{ij}) \geq h_b$ **then**

18: Add edge e_{ij} to Ω_b

19: **end if**

20: **end for**

21: $\Omega = \Omega_b \cup \tilde{\Omega}$.

22: **return** Ω

Computational complexity

In this subsection, we will give the computational complexity of the two motif-based methods: one is the Motif-DF (Algorithm 1), the other one is our newly proposed Motif- h

Fig. 2 The process of extracting the motif h -backbone of the food web network. The subgraphs **a–f** in left side consist of edges with strength larger than or equal to the motif h -strength or h -bridge accordingly of the food web network; **g** a illustration of the original food web network. The nodes with darker color have larger degrees; **h** the connected motif h -backbone of the food web network

(Algorithm 2). The computational complexity of two filter-based methods are mainly determined by the two stages: (a) the edges' weights computing and (b) extracting the backbone in the network $\mathcal{G}$.

In Motif-DF method, the complexity of computing $\hat{w}_{ij}$ is $\mathcal{O}(|\mathcal{E}| \cdot \bar{k})$ and the computing τ_{ij} is $\mathcal{O}(|\mathcal{N}| \cdot \bar{k})$, where $|\mathcal{E}|$ is the number of edges and $\bar{k}$ is the average degree in network $\mathcal{G}$. In our proposed Motif- h , the time of computing h -strength is $\mathcal{O}(|\mathcal{E}|)$, and time of computing h -bridge is $\mathcal{O}(|\mathcal{E}| \cdot \mathcal{N})$. Both complexity of extraction processes are $\mathcal{O}(|\mathcal{E}|)$.

Evaluation metrics

Unlike the existing backbone problem, a functional backbone extraction issue aims to preserve as much as possible the salient motifs in the original network.

We choose the food web of Mangrove Estuary in dry seasons as an example. The network has 97 vertices and 1491 edges, where nodes represent species and edge represent energy flow among in the community [10]. In the food

web network, the salient motifs are $M_2, M_3, (M_8 - M_{10})$, in which M_2 is the most significant motif. As shown in Fig. 2, these salient motifs represent the natural properties of the network; when looking for the network skeleton, the final backbone should hold as many salient motifs as possible.

To evaluate the Motifs reserved in the functional backbone, we adopt the motif degree and motif centrality as two importance metrics in the experiments:

Definition 5 *Motif Degree*: Given a salient motif M_k , the motif degree d_{M_k} of M_k is defined as the sum of nodes' degree in M_k , and the total motif degree is defined as $MD = \sum_{M_k \in \mathbb{M}} d_{M_k}$ in the current $\mathcal{G}$.

Definition 6 *Motif Centrality*: Given a salient motif M_k , the motif centrality is defined as the average centrality of all its occurrences in the network. Suppose the total number of occurrence of M_k in $\mathcal{G}$ is n , the motif centrality of M_k can be calculated as follows:

$$MC(M_k) = \frac{1}{(n-1)(n-2)} \sum_{i=1}^l \sum_{j=1}^m BC(v_j)/m, \quad (16)$$

where $BC(v_j)$ is the betweenness of some node v_j , and m is the number of nodes in the motif, in this paper, we use $m = 3$. The total motif centrality in $\mathcal{G}$ is

$$MC(\mathbb{M}) = \sum_{M_k \in \mathbb{M}} MC(M_k). \quad (17)$$

Experiments and results

In this section, we conduct experiments to validate the effectiveness of our Motif- h compared with Motif-DF and Entropy methods under four real-world networks. The details of the data are as follows.

- The Open flight network: The network we obtain is from the OpenFlights/Airline Route Mapper Route Database. It contains more than 60,000 routes between 3425 airports on 548 airlines around the world. We transform the data into a directed network with 3425 nodes and 37,595 edges after removing duplicated edges [24]. The data are available at <https://openflights.org/data.html>.
- The WikiElec is a voting network in which users can support (trust) or oppose (distrust) other users in administrator elections, which consists of 7194 nodes and 103,591 edges [50]. The data are available at <https://github.com/WHU-SNA/STNE/tree/master/input>.
- The Transportation reachability network: It is a network of reachable cities in the United States and Canada, which includes 71,959 traffic paths between 456 cities [15,27]. The data are available at <http://snap.stanford.edu/data>.
- The Facebook (NIPS) is social networks belong to the conference, which includes 2888 people and 2981 connections. The data are available at <http://www.konect.cc>.

Before the backbone extraction procedures, we firstly calculate the SP values of each network to obtain the corresponding salient motifs set and the results are shown in Table 2. Obviously, there is variability in the kinds of salient motifs for different types of networks, which means that the frequency of occurrence of specific motif exposes the essential characteristics of the network.

In simulations, we compare our proposed Motif- h with two state-of-the-art, Motif-DF and Entropy methods, in terms of the motif degree and motif centrality preserving. For a fair comparison, we compare the backbone extraction of all nodes vs the two metrics. As the extraction of Motif- h is finished, we will extract the remaining nodes, randomly. Besides, Entropy method is a representative global threshold

Table 2 The salient ID in the real networks: an illustration of the $\mathcal{SP}$ values for all 13 three-node motifs of four real-world networks

Networks	Salient Motif
Open flight	$M_1, M_2, M_3, M_4, M_5, M_6, M_7$
WikiElec	$M_1, M_2, M_5, M_7, M_8, M_9$
Reachability network	$M_4, M_6, M_7, M_9, M_{11}, M_{12}$
Facebook (NIPS)	$M_3, M_4, M_8, M_{10}, M_{11}, M_{12}$

Motifs with $\mathcal{SP}$ values larger than zero are recognized as salient motifs. Among the four real-world networks, the salient motifs are quite different

method, and the corresponding edge weights are computed as

$$\hat{w}_{ij}^E = \hat{w}_{ij} \log \hat{w}_{ij}. \quad (18)$$

Moreover, trust networks are the special kind of directed networks [23], which is very common in social networks, and represents trust when the weight of the edge is greater than 0 [1], and vice versa for distrust. In this paper, we focus on the WikiElec network motif backbone problem, which is a common trust network [49]. Actually, the trust network extraction is not a simple task, often losing untrusted edges because of the comparison of edge real value. In the simulation comparisons, we split the trust network into two sub-networks $\mathcal{G} = \mathcal{G}^+ \cup \mathcal{G}^-$, containing the network's positive and negative edges, respectively.

With the initial selected salient motif set M according to Eq. (3), we calculate the bridging values and the motif edge strengths of all edges in each salient motif M_k . Then filter the candidate edges with $\tilde{w}_{ij} \geq h_{\mathbb{M}}$ and $\mathcal{B}(e_{ij}) \geq h_b$, according to the motif edge strength values and bridging values, respectively.

By merging the edges qualified in the Ω_b and $\tilde{\Omega}$, we achieve the final motif h -backbone set Ω . We show the procedures of extracting the motif h -backbone in Fig. 2. The left side of Fig. 2 shows the process of extracting the motif h -backbone of the food web network, panel (g) is the original food web network and the panel (h) is its motif h -backbone. Clearly, the original network is too dense to get useful information visually. Our Motif- h gives an abstract architecture by extracting the subgraphs from the dense network. The subgraphs on the left side of Fig. 2 are much more straightforward than the original network, which displays the edges coming from each salient motifs. The edges in the backbone that come from M_{10} are in the majority, which describes respirations in the energy flow in the food web network. By combining the edges in the subgraphs, we obtain the functional backbone of the whole food web network.

Fig. 3 The changes of Motif degree after removing a proportion of backbone from the current networks. **a** The Open flight network; **b** the Facebook (NIPS) networks; **c** the Transportation reachability network; **d** the WikiElec network. The horizontal axis of the graph indicates the proportion of edges removed. The vertical axis of the graph represents the network degree after removing corresponding edges.

Comparisons of motif degree and centrality

To verify the proposed method's effectiveness in this paper, we have conducted comparative experiments on the three methods, i.e., Motif-*h*, Motif-DF, and Entropy, under four real-world networks. The network motif degree is an essential measure of the functional network backbone. For the detailed comparison, we recorded the remaining network after removing the important edges. The faster the network motif degree decreases, the more critical the extracted backbone is. As shown in Fig. 3 shows the whole process from the complete network to zero motif degree is recorded with an interval of 2%.

It is visible that our proposed algorithm decreases the motif degree the fastest among the four types of networks and requires the least number of skeletons to reduce the motif degree of all networks. These results indicate that our motif-*h*

has a clear advantage in selecting the backbone and finding the more important network core structures with limited edges.

Note that three methods obtain similar results in the Facebook network, mainly because the number of edges in this network is not dense. As the network tends to be sparse, most exist edges in the networks tend to be necessary, and the sequence of edge selections in different methods becomes more consistent.

Figure 4 shows the comparisons in terms of motif centrality. Similar to the results on motif degree, we found that our proposed Motif-*h* method has a significant priority over Motif-DF and Entropy. Using motif-*h* to find the functional backbone, in four real-world networks, we only need 30% of extracted edges to reduce the motif centrality in the original networks significantly.

Fig. 4 The changes of Motif centrality after removing a proportion of backbone from the current networks. **a** The Open flight network; **b** the Facebook(NIPS) networks; **c** the transportation reachability network; **d** the WikiElec network. The horizontal axis of the graph indicates the proportion of edges removed. The vertical axis of the graph represents the network centrality after removing corresponding edges

Fig. 5 The thresholds of different methods as the motif measures dropping to zeros. The gray bar represents the Entropy method, the orange bar represents the Motif-DF method and the blue bar represents our method.

In addition to the two important indicators' results, we also specifically analyze the minimum backbone ratios required when the two indicators' values turn to almost zeros. Comparing with the two sub-graphs in Fig. 5, we found different algorithms have similar thresholds in both cases, Motif

degree and Motif centrality. Comprehensively observing the three algorithms' actual performance, we can notice that the thresholds of the motif proposed in this article are always in the lowest state, which validates our method's effectiveness.

Conclusion

Unlike traditional backbone extraction methods, we motivate a new direction to find the challenging functional backbone to preserve the high-order structures, *network motifs*. In this work, we first introduce the Motif-*h* backbone extraction method based on the *h*-index and bridge nodes, to do such remarkable backbone findings. The new method utilizes the *h*-index measure to understand the individual network better and avoids certain global thresholds, which are two critical limitations often found in the most existing methods. Besides, this paper focuses on the vital role of weak connectivity in the network, and we first incorporate it into the backbone algorithm design, resulting in a unique two-stage selection mechanism. Simulation results show that our proposed Motif-*h* have much better performance in vital backbone finding and motifs preserving than two important methods, i.e., Motif-DF and Entropy.

Different types of networks often have typical high-order structures, and it is significant to use high-order structures to define the network skeleton. The research on module-based network backbone search has just begun, and many problems still need to be solved: how to efficiently find network motifs? [2,47] How to forward the motifs results to the special network, e.g., temporal networks? [26,36]

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant nos. 11801430, 61877046, and 61877047), the Natural Science Basic Research Program of Shaanxi (Grant nos. 2020JM-178 and 2021JM-115), the Fundamental Research Funds for the Central Universities (Grant no. YJS2107), and the Innovation Fund of Xidian University. Y.G. Bai thanks the China Scholarship Council (CSC) for financial support.

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/>.

References

1. Aghdam NH, Ashtiani M, Azgomi MA (2020) An uncertainty-aware computational trust model considering the co-existence of trust and distrust in social networks. *Inf Sci* 513:465–503
2. Al-Thaadan, A., Carvalho, M.: Online estimation of motif distribution in dynamic networks. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, pp 0758–0764 (2019)
3. Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8(6):450–461
4. Bai Y, Liu S, Li Q, Yuan J (2020) Cost-aware deployment of check-in nodes in complex networks. *IEEE Trans Syst Man Cybern Syst.* <https://doi.org/10.1109/TSMC.2020.3034485>
5. Bai Y, Liu S, Zhang Z (2017) Effective hybrid link-adding strategy to enhance network transport efficiency for scale-free networks. *Int J Mod Phys C* 28(08):1750107
6. Bai Y, Yuan J, Liu S, Yin K (2019) Variational community partition with novel network structure centrality prior. *Appl Math Model* 75:333–348
7. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5(2):101–113
8. Bascompte J (2009) Disentangling the web of life. *Science* 325(5939):416–419
9. Bassett DS, Sporns O (2017) Network neuroscience. *Nat Neurosci* 20(3):353–364
10. Batagelj V, Mrvar A (2006) Pajek datasets. <http://vlado.fmf.uni-lj.si/pub/networks/data/>. Accessed on 24 feb 2021
11. Benson AR, Gleich DF, Leskovec J (2016) Higher-order organization of complex networks. *Science* 353(6295):163–166
12. Cao J, Ding C, Shi B (2019) Motif-based functional backbone extraction of complex networks. *Phys A Stat Mech Appl* 526:121123
13. Chen X (2015) Critical nodes identification in complex systems. *Complex Intell Syst* 1(1–4):37–56
14. Coscia M, Neffke FM (2017) Network backboning with noisy data. In: 2017 IEEE 33rd International Conference on Data Engineering (ICDE). IEEE, pp 425–436
15. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. *Science* 315(5814):972–976
16. Gfeller D, De Los Rios P (2007) Spectral coarse graining of complex networks. *Phys Rev Lett* 99(3):038701
17. Gfeller D, De Los Rios P (2008) Spectral coarse graining and synchronization in oscillator networks. *Phys Rev Lett* 100(17):174104
18. Ghalmane Z, Cherifi C, Cherifi H, El Hassouni M (2020) Extracting backbones in weighted modular complex networks. *Sci Rep* 10(1):1–18
19. Gong Y, Liu S, Bai Y (2021) Efficient parallel computing on the game theory-aware robust influence maximization problem. *Knowl Based Syst* 220:106942
20. Grady D, Thiemann C, Brockmann D (2012) Robust classification of salient links in complex networks. *Nat Commun* 3(1):1–10
21. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
22. Honey CJ, Kötter R, Breakspear M, Sporns O (2007) Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc Natl Acad Sci* 104(24):10240–10245
23. Jiang C, Liu S, Lin Z, Zhao G, Duan R, Liang K (2016) Domain-aware trust network extraction for trust propagation in large-scale heterogeneous trust networks. *Knowl Based Syst* 111:237–247
24. OpenFlights, O. F. (2017). Airport, airline and route data. 2017 <http://openflights.org/data.html>
25. Kim DH, Noh JD, Jeong H (2004) Scale-free trees: the skeletons of complex networks. *Phys Rev E* 70(4):046126
26. Kosyfaki C, Mamoulis N, Pitoura E, Tsaparas P (2018) Flow motifs in interaction networks. *arXiv preprint arXiv:1810.08408*
27. Leskovec J, Krevl A (2014) SNAP datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>. Accessed on Jun 2021
28. Lin L, Wu C, Ma L. A genetic algorithm for the fuzzy shortest path problem in a fuzzy network[J]. *Complex & Intelligent Systems*, 2021, 7(1): 225–234
29. Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63

30. Malang K, Wang S, Lv Y, Phaphuangwittayakul A (2020) Skeleton network extraction and analysis on bicycle sharing networks. *Int J Data Warehous Min (IJDWM)* 16(3):146–167
31. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci* 100(21):11980–11985
32. Märtens M, Meier J, Hillebrand A, Tewarie P, Van Mieghem P (2017) Brain network clustering with information flow motifs. *Appl Netw Sci* 2(1):25
33. McDonnell MD, Yaveroğlu ÖN, Schmerl BA, Iannella N, Ward LM (2014) Motif-role-fingerprints: the building-blocks of motifs, clustering-coefficients and transitivity in directed networks. *PLoS One* 9(12):e114503
34. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U (2004) Superfamilies of evolved and designed networks. *Science* 303(5663):1538–1542
35. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. *Science* 298(5594):824–827
36. Paranjape A, Benson AR, Leskovec J (2017) Motifs in temporal networks. In: *Proceedings of the tenth ACM international conference on web search and data mining*. Cambridge, United Kingdom, pp 601–610
37. Phaphuangwittayakul A (2018) From complex network to skeleton: MJ-modified topology potential for node importance identification. In: *Proceedings of advanced data mining and applications: 14th international conference*, vol 11323, ADMA 2018, Nanjing, China, November 16–18, 2018. Springer, pp 413
38. Purohit S, Holder LB, Chin G (2020) Item: independent temporal motifs to summarize and compare temporal networks. *arXiv preprint arXiv:2002.08312*
39. Radicchi F, Ramasco JJ, Fortunato S (2011) Information filtering in complex weighted networks. *Phys Rev E* 83(4):046101
40. Rosvall M, Esquivel AV, Lancichinetti A, West JD, Lambiotte R (2014) Memory in network flows and its effects on spreading dynamics and community detection. *Nat Commun* 5(1):1–13
41. Saramäki J, Kivela M, Onnela JP, Kaski K, Kertesz J (2007) Generalizations of the clustering coefficient to weighted complex networks. *Phys Rev E* 75(2):027105
42. Serrano MÁ, Boguná M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proc Natl Acad Sci* 106(16):6483–6488
43. Song C, Havlin S, Makse HA (2005) Self-similarity of complex networks. *Nature* 433(7024):392–395
44. Wang JW, Rong LL (2009) Cascade-based attack vulnerability on the us power grid. *Saf Sci* 47(10):1332–1336
45. Wang P, Lü J, Yu X (2014) Identification of important nodes in directed biological networks: a network motif approach. *PLoS One* 9(8):e106132
46. Wang S, Malang K, Yuan H, Phaphuangwittayakul A, Lv Y, Lowdermilk MD, Geng J (2020) "Extracting Skeleton of the Global Terrorism Network Based on m-Modified Topology Potential", *Complexity*, vol. 2020, Article ID 7643290, 18 pages, 2020
47. Wernicke S (2006) Efficient detection of network motifs. *IEEE/ACM Trans Comput Biol Bioinform* 3(4):347–359
48. Williams RJ, Martinez ND (2000) Simple rules yield complex food webs. *Nature* 404(6774):180–183
49. Xu P, Hu W, Wu J, Liu W (2020) Opinion maximization in social trust networks. *arXiv preprint arXiv:2006.10961*
50. Xu P, Hu W, Wu J, Liu W, Du B, Yang J (2019) Social trust network embedding. In: *2019 IEEE international conference on data mining (ICDM)*. Beijing, China, pp 678–687
51. Yuan H, Han Y, Cai N, An W (2018) A multi-granularity backbone network extraction method based on the topology potential. *Complexity*
52. Zeng A, Lü L (2011) Coarse graining for synchronization in directed networks. *Phys Rev E* 83(5):056123
53. Zeng L, Jia Z, Wang Y (2019) A new spectral coarse-graining algorithm based on k-means clustering in complex networks. *Mod Phys Lett B* 33(01):1850421
54. Zhang RJ, Stanley HE, Fred YY (2018) Extracting h-backbone as a core structure in weighted networks. *Sci Rep* 8(1):1–7
55. Zhang X, Zhang Z, Zhao H, Wang Q, Zhu J (2014) Extracting the globally and locally adaptive backbone of complex networks. *PLoS One* 9(6):e100428
56. Zhang X, Zhu J (2013) Skeleton of weighted social network. *Phys A Stat Mech Appl* 392(6):1547–1556
57. Zhao C, Bin A, Ye W, Fan Y, Di Z (2015) Motif for controllable toggle switch in gene regulatory networks. *Phys A Stat Mech Appl* 419:498–505
58. Zhao SX, Zhang PL, Li J, Tan AM, Ye FY (2014) Abstracting the core subnet of weighted networks based on link strengths. *J Assoc Inf Sci Technol* 65(5):984–994

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