# Motif-Aware Diffusion Network Inference

Qi Tan[1], Yang Liu[1,2], and Jiming Liu[1,2](✉)

[1] Department of Computer Science, Hong Kong Baptist University, Hong Kong, China
{csqtan,csygliu,jiming}@comp.hkbu.edu.hk
[2] IRACE Research Center, Hong Kong Baptist University, Shenzhen, China

**Abstract.** Characterizing and understanding information diffusion over social networks play an important role in various real-world applications. In many scenarios, however, only the states of nodes can be observed while the underlying diffusion networks are unknown. Many methods have therefore been proposed to infer the underlying networks based on node observations. To enhance the inference performance, structural priors of the networks, such as sparsity, scale-free, and community structures, are often incorporated into the learning procedure. As the building blocks of networks, network motifs occur frequently in many social networks, and play an essential role in describing the network structures and functionalities. However, to the best of our knowledge, no existing work exploits this kind of structural primitives in diffusion network inference. In order to address this unexplored yet important issue, in this paper, we propose a novel framework called Motif-Aware Diffusion Network Inference (MADNI), which aims to mine the motif profile from the node observations and infer the underlying network based on the mined motif profile. The mined motif profile and the inferred network are alternately refined until the learning procedure converges. Extensive experiments on both synthetic and real-world datasets validate the effectiveness of the proposed framework.

## 1 Introduction

Characterizing and understanding information diffusion processes over social networks play an important role in many real-world applications, such as viral marketing [10] and rumor detection [3]. However, in many scenarios, the underlying diffusion networks are hidden [19,20]; what we do have is the states of nodes observed over time. Therefore, inferring the underlying networks based on the observations of node states is of great importance and has received much attention recently [5,19,20].

Utilizing the network structure properties (e.g., community structure [8] and scale-free property [21]) as the prior in the inference procedure has been proved effective in improving the performance of network inference [7,18]. Network motifs, which are regarded as the building blocks of networks [1,17], occur frequently in many real-world networks and play a key role in analyzing the network structure and interpreting the network functionality. For example, as an


D. Phung et al. (Eds.): PAKDD 2018, LNAI 10939, pp. 638–650, 2018.
https://doi.org/10.1007/978-3-319-93040-4_50

important mesoscale structure, motif patterns characterize the local structure of the network connectivity and contribute to network classification [16] and community detection [2]. Moreover, the incoherent feed-forward loop, which is a representative triangle network motif, is commonly found in gene regulation networks and can provide fold-change detection [4].

Although motifs are of great importance in describing the network structures and functionalities, to the best of our knowledge, no existing work exploits this kind of structural primitives in diffusion network inference. In order to address this unexplored yet important issue, we propose a novel framework called Motif-Aware Diffusion Network Inference (MADNI), which takes the network motifs into account when inferring the underlying diffusion networks. Figure 1 schematically illustrates the idea and procedure of the proposed framework.

The contributions of this paper are summarized as follows.

1. We investigate an unexplored yet important issue, i.e., how to integrate motif prior into diffusion network inference.
2. We propose a novel learning framework MADNI to mine the motif profile and incorporate the uncovered motif profile into the network inference procedure.
3. We perform extensive experiments on both synthetic and real-world datasets, showing the effectiveness of the proposed framework.

## 2 Related Work

The proposed MADNI aims to jointly mine the network motif prior and infer the underlying diffusion network. This section therefore reviews some related works in underlying diffusion network inference and the inference with network priors.

The diffusion network inference problem refers to tracing the diffusion edges based on the observed infection time sequence. Gomez et al. proposed an algorithm NETINF [5] to infer the diffusion edges through maximizing the likelihood of observed infection time by utilizing submodular optimization. To infer the heterogeneous transmission rates and time-varying network, NETRATE [19] and INFOPATH [6] have been proposed respectively. Rong et al. proposed a model-free approach NPDC [20] to utilizes the statistical difference of the infection time intervals between nodes connected with diffusion edges versus those without diffusion edges in network inference. Moreover, Hu et al. proposed a clustering embedded approach CENI [9] to improve the efficiency of network inference by clustering the nodes on the embedded space.

Incorporating the network prior into the learning procedure generally improves the performance of network inference [8,13,21]. Many literatures have focused on inferring the scale-free networks and modular networks with block structure. Liu and Ihler added a log $l_1$ norm regularization on the estimated graph structure in the Gaussian graphical model learning to encourage the estimated graph becoming scale-free [15]. Liu et al. introduced the weight inverse graph prior to encourage specific node distribution [13]. Hosseini Lee. introduced a block prior to encourage sparse connections between blocks [8].

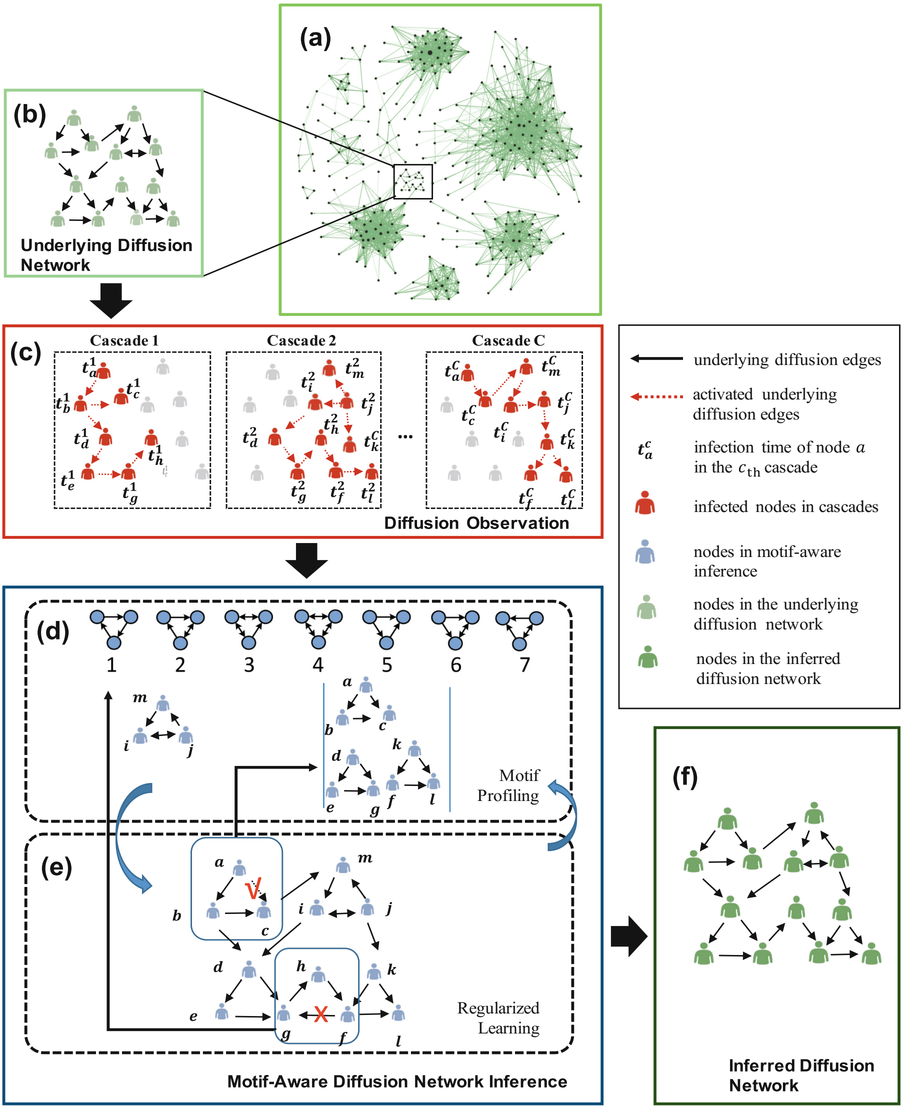


**Fig. 1.** Schematic illustration of the proposed motif-aware diffusion network inference (MADNI). (a) An example of the diffusion network. (b) An enlarged part of the diffusion network in (a) for algorithm illustration. (c) The observed cascades on the underlying diffusion network shown in (b). For each cascade, only the infection time of the influenced nodes (the red nodes) are observed, such as $\mathbf{t}^1 = \{t_a^1, t_b^1, \cdots\}$. (d) The motif profile is mined from the cascade data by estimating the frequency of various motif patterns in the underlying diffusion network. (e) The underlying diffusion network is learned via motif prior regularized learning. The mined motif profile and the learned network are alternately refined until the inferred network converges. (f) The diffusion network inferred by the MADNI framework. (Color figure online)

However, motif, the structural primitives of many real-world networks, has not been taken into account in underlying diffusion network inference yet. Therefore, in this paper, we attempt to fill this gap by proposing a novel framework to discover the underlying motif patterns and incorporate the uncovered motif patterns into the network inference procedure.

## 3 Motif-Aware Diffusion Network Inference

In this section, we present our framework, Motif-Aware Diffusion Network Inference (MADNI). First, we provide the necessary notations and formally define the problem. Then we introduce how to estimate the initial structure motif profile from the cascade data and present a novel scheme for inferring the diffusion network with the mined structural motif profile.

### 3.1 Notations and Problem Formulation

Let $G =< V, E >$ denote the directed diffusion network, where $V$ denotes the set of $N$ nodes (representing the individuals, Blog sites or locations) and $E$ denotes the set of edges (representing the directed influence from one node to another). Generally, the edges in $E$ is represented by the $N \times N$ adjacency matrix $A$, where an entry (i, j) of $A$, $A_{ij}$, is the transmission rate from node $i$ to node $j$. A subgraph $G_s = (V_s, E_s)$ of $G$ satisfies $V_s \subseteq V$, $E_s \subseteq E$. Two subgraphs $G_s^1 = (V_1, E_1)$ and $G_s^2 = (V_2, E_2)$ are isomorphic if there exists a projection $\varphi : V_1 \rightarrow V_2$ with $(u, v) \in E_1 \leftrightarrow (\varphi(u), \varphi(v)) \in E_2$ for all $(u, v)$.

**Motif**. A motif pattern with $k$ nodes is a non-isomorphic, connected subgraph frequently appearing in a large network. Figure 1(d) shows all seven close connected triangle motif patterns.

Consider that the diffusion observation $O$ is collected over the underlying diffusion network $G$ and consists of a set of $C$ cascades. Each cascade $t^c$ ($c = 1, ..., C$) is a collection of observed infection time stamps within the population during a time interval of length $T$ and can be represented as an $N$-dimensional vector $t^c := (t_1^c, \cdots, t_N^c)$, where $t_n^c \in [0, T] \cup \{\infty\}$ indicates the infection time of node $n$ in cascade $c$. The symbol $\infty$ labels users that are not infected during observation window $[0, T]$. Given the above diffusion observation $O$, we aim to infer the underlying relation between nodes on $G$, i.e., the adjacency matrix $A$.

### 3.2 Estimating Motif Pattern from Cascade Data

In this subsection, we introduce a straightforward yet effective approach to estimate the motif frequency from the cascade data, which is expected to be helpful in inferring the underlying diffusion network. When scanning the cascade sequences, we record the occurrence matrix $CO \in R^{N \times N}$ as follows. In each cascade, if $t_i^c + t^W > t_j^c > t_i^c$, where $t^W$ is the time window, $CO_{i,j}$ increases by one. Therefore, $CO_{i,j}$ reflects how many times that node $i$ may influence node

$j$. We could further take into account the interval between the infection time of node $i$ and that of node $j$, i.e., $CO_{i,j}$ increases by $\sigma(-(t_j^c - t_i^c))$, where $\sigma(\cdot)$ could be the Exponential function or Rayleigh function [19]. Based on this occurrence matrix, we extract the significant pairwise influences by calculating the score of the edge count: $S_{i,j}^E = \frac{CO_{i,j} - mean(CO)}{std(CO)}$, where $mean(CO)$ and $std(CO)$ denote the mean and standard deviation of all elements in $CO$, respectively. The edges with low scores are filtered out. Then we could count the subgraph frequency $f^m$ based on the $S^E$ by assuming that there exists an edge $e_{i,j}$ if $S_{i,j}^E$ is nonzero, and further determine the significant motifs by calculating the $Z$-score of the subgraph as $Q^m = \frac{f^m - mean(\mathcal{F}^m)}{std(\mathcal{F}^m)}$, where $\mathcal{F}^m$ is the $m$-th motif frequency of a set of samples drawn by randomly shuffling $S^E$ [17]. The $m$-th motif is significant if $Z^m$ is far above 1 [16]. We denote the procedure of extracting significant motifs as $\mathcal{M} \leftarrow \mathcal{Q}(S^E)$.

### 3.3 Motif Prior Regularization

Different types of networks exhibit distinct motif frequency profiles [16]. In order to incorporate the motif prior into network inference, we propose an edge-centered regularization to adjust the motif frequency profile of the estimated network. The main idea is that if an edge is forming a high frequent motif, less penalization will be given to it. Let $Z^m \in \Re^{N \times N}$ denote the motif count matrix of $G$ for a certain motif pattern $m$, where $Z_{i,j}^m$ indicates the number of instances of motif $m$ that containing the edge $(i \rightarrow j)$ [2]. We detect the motif $m$ from the adjacency matrix $A$ and count $Z_{i,j}^m$ for each edge, which is denoted as $Z^m = \mathcal{P}(A, m)$. Furthermore, we have $Z = \sum_{m \in \mathcal{M}} Z^m = \sum_{m \in \mathcal{M}} \mathcal{P}(A, m) = \mathcal{P}(A)$. The motif pattern in $\mathcal{M}$ could be selected as significant motifs detected from the networks or based on prior knowledge. In this paper, the seven close connected triangle motif patterns shown in Fig. 1 are considered since the empirical studies have revealed that these motif patterns appear frequently and play special roles in social networks [14]. Based on the motif count matrix, the reweighted regularization can be constructed for learning the network with significant frequent motif patterns:

$$R(\mathcal{P}(A)) = |M \circ A| = \sum_{i,j=1}^{N} |\frac{A_{i,j}}{Z_{i,j} + 1}| \tag{1}$$

### 3.4 Learning

We aim to find the diffusion network such that likelihood of diffusion observation is maximized. The likelihood of diffusion observation is calculated as follow.

**Pairwise transmission likelihood.** With the cascades data, the pairwise transmission likelihood is calculated as follows. Define $f(t_i^c | t_j^c, A_{j,i})$ as the transmission likelihood from node $j$ to node $i$, which is related to the infection time interval $\triangle t = (t_j^c - t_i^c)$ and the transmission rate $A_{j,i}$. Moreover, a node can only be infected by an infected node. The exponential parametric likelihood model is adopted: $f(t_i^c|t_j^c, A_{j,i}) = A_{j,i} \exp^{-A_{j,i}(t_i^c - t_j^c)}$ if $t_j^c < t_i^c$ and 0 otherwise. The survival likelihood of edge $(j \rightarrow i)$, denoted as $S(t_i^c|t_j^c, A_{j,i})$, is the probability that node $i$ is not infected by node $j$ by time $t_i^c$, which is calculated as: $S(t_i^c|t_j^c, A_{j,i}) = 1 - F(t_i^c|t_j^c, A_{j,i})$ , where $F(t_i^c|t_j^c, A_{j,i})$ is the cumulative function of the transmission likelihood.

**Likelihood of a cascade.** The likelihood of the observe infections $\hat{t}^c = (t_1^c, \cdots, t_N^c)$ is calculated as:

$$f(\hat{t}^c; A) = \prod_{t_i^c \leq T} \prod_{t_m^c > T} S(T|t_i^c, A_{i,m}) \prod_{k:t_k^c < t_i^c} S(t_i^c|t_k^c, A_{k,i}) \times \sum_{j:t_j^c < t_i^c} H(t_i^c|t_j^c, A_{j,i}), \tag{2}$$

where $H(t_i^c|t_j^c, A_{j,i})$ is the hazard function: $H(t_i^c|t_j^c, A_{j,i}) = \frac{f(t_i^c|t_j^c, A_{j,i})}{S(t_i^c|t_j^c, A_{j,i})}$.

**Network inference.** We aim to search $A$ that maximizes the likelihood of cascade observation $O$. In our framework, the networks are estimated through maximizing the regularized likelihood function as follow.

$$\begin{aligned} &\max_A \Big( L(O|A) - R(\mathcal{P}(A)) \Big) = \max_A \Big( \sum_{c \in C} \log f(t^c, A) - R(\mathcal{P}(A)) \Big) \\ &s.t.\ A_{j,i} \geq 0, i, j = 1, \cdots, N \end{aligned} \tag{3}$$

where $L(O|A)$ is the likelihood function of observation given the network topology and $R(\mathcal{P}(A))$ is the regularization term. We can use ADMM or projected gradient descent [6] to enforce $A$ to be nonnegative. Thus the matrix gradient in terms of $A$ is written as $\frac{\partial L}{\partial A} - M$. The additional computational bounden of adding the regularization term is just the addition of a $N \times N$ matrix. The gradient for edges linking to node $k$ in the cascade where node $k$ is uninfected is

$$\frac{\partial L^c}{\partial A_{j,k}} = T - t_j^c \tag{4}$$

and the gradient for edges linking to node $k$ in the cascades where node $k$ is infected is:

$$\frac{\partial L^c}{\partial A_{j,k}} = (t_k^c - t_j^c) - \frac{1}{\sum_{l:t_l^c < t_k^c} A_{l,k}} \tag{5}$$

Summating the above term over all cascades gives the gradient for edges linking to node $k$. Starting from $Z = \mathcal{P}(CO)$ or a plain prior, i.e., $Z = \mathcal{P}(\mathbf{0}^{N \times N})$, we update the network structure and the motif count matrix alternately until the estimated network structure remains unchanged. The detailed procedure of the proposed MADNI framework is provided in Algorithm 1.

### 3.5 Computational Complexity Analysis

In this subsection, we analyze the computational complexity of the proposed framework. The time cost of occurrence matrix counting is $O(CL)$, where $C$ is the

**Algorithm 1:** Motif-Aware Diffusion Network Inference (MADNI)

**Input** : The observation $O$: $C$ cascades $\{t^c = (t_1^c, ..., t_N^c)|_{c=1}^C\}$
**Output**: The estimated network $\hat{A}$
**for** $c = 1, \cdots, C$ **do**
  **if** $t_i^c + t^W > t_j^c > t_i^c$ **then**
    $CO_{i,j} \leftarrow CO_{i,j} + \sigma(-(t_j^c - t_i^c))$; ▷ `Construct occurrence matrix`
  **end**
**end**
**for** $i, j = 1, \cdots, N$ **do**
  $S_{i,j}^E \leftarrow \frac{CO_{i,j} - mean(CO)}{std(CO)}$ ; ▷ `Calculate edge significance`
**end**
$\mathcal{M} \leftarrow \mathcal{Q}(S^E)$ ; ▷ `Initialize candidature motifs set` $\mathcal{M}$
$Z \leftarrow \mathcal{P}(S^E, \mathcal{M})$ ; ▷ `Initialize motif profile`
**while** *not converged* **do**
  $\hat{A} \leftarrow \arg\max_A L(O|A) - R(Z)$ ; ▷ `Learn diffusion network`
  $Z \leftarrow \mathcal{P}(\hat{A}, \mathcal{M})$ ; ▷ `Update motif profile`
**end**

number of cascades and $L$ is the length of a cascade. Assume on average there are $K$ elements in each row of $S^E + (S^E)^T$, then the cost of motif counting in $S^E$ and its random shuffling variants is $O(NK^2)$. For each iteration of diffusion network learning and motif profile updating, the computational demand comes from two parts: motif count matrix calculation and network inference. Assume there are $M$ edges in the graph and the maximum degree is $D_{max}$. For each edge $e$, the cost of calculating the motif count $C_e$ is $O(D_{max})$ and thus the cost of calculating the matrix count for all edges is $O(MD_{max})$. The network structure is inferred via an iterative way, in which the complexity in each iteration is $O(CN^2)$. If the maximum number of iterations in network inference problem is $N_i$, then the complexity of network inference is $O(N_iCN^2)$. As $O(CN_iN^2) > O(MD_{max})$ holds in general, the total computational cost of the proposed framework is $O(CN_iN^2)$.

## 4 Validations

In this section, we evaluate the performance of our framework in diffusion network inference on both the synthetic and real-world networks, in terms of Precision, Recall and $F1$ score [7].

### 4.1 Experiments on Synthetic Networks

In this subsection, we evaluate the performance of our framework on synthetic networks and cascades. We first construct the synthetic network with $N$ nodes and then generate $C$ cascades using exponential diffusion model on the network

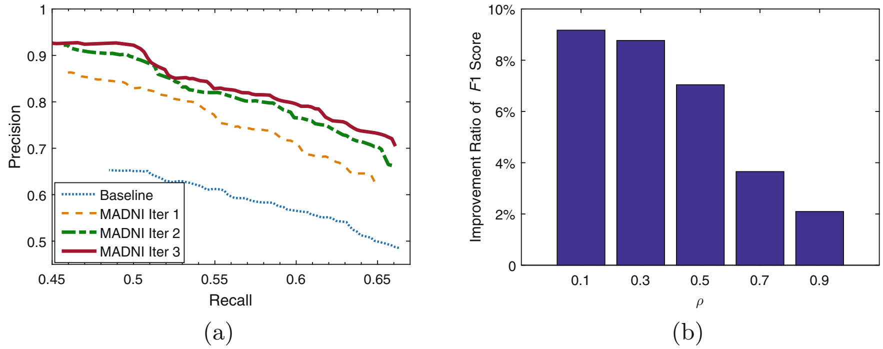


**Fig. 2.** (a) Performance comparison (in terms of Precision and Recall) between MADNI and the baseline. MADNI significantly outperforms the baseline method with only one iteration. The performance of MADNI is further improved after each iteration, and the learning procedure quickly converges in only three iterations. (b) $F1$ improvement ratio of MADNI over the baseline method with varying $\rho$. MADNI outperforms the baseline even if the target network is close to a random network ($\rho = 0.9$). When the target network becomes more structured, i.e., $\rho$ becomes smaller, the improvement ratio becomes more significant.

as the observation [19]. We choose NETRATE [19] as the baseline method for comparison in our experiments as it explores the global convexity of network inference problem.

**Comparison with baseline**. We first evaluate the performance of the proposed framework on inferring the motif-dense network, i.e., the network with certain motif patterns occurring frequently. The motif-dense network is generated from a random motif network model, which enumerates the combinations over all nodes and assigns a specific motif to each node combination with probability $p$. Here we choose the feed-forward loop motif, i.e., Motif 5 illustrated in Fig. 1(d), for our experiment, as it is a commonly observed motif in social networks [16]. Figure 2(a) shows the Precision and Recall of the baseline method and those of MADNI after different number of iterations. It can be seen that the proposed framework achieves significant improvement over the baseline method with only one iteration, which validates the effectiveness of taking the motif into consideration when inferring the structured networks. Furthermore, the performance of the proposed framework can be further improved after each iteration, and the learning procedure quickly converges in only three iterations.

**Performance improvement with random edges**. After having evaluated the performance of the proposed framework on the motif-dense network, we further test our framework on the networks consist of both significant motifs and random edges. Specifically, we generate the target network from a motif-dense network and a random network. The proportion of random network is indicated by a parameter $\rho$, where $\rho = 0$ indicates the complete motif-dense network, which is used in our previous experiment; while $\rho = 1$ indicates the complete random

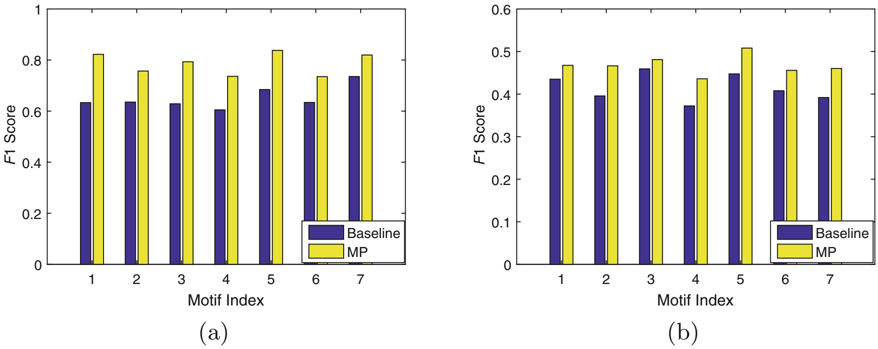


**Fig. 3.** Performance comparison (in terms of $F1$ score) between MADNI and the baseline in (a) Exponential and (b) Rayleigh cascades on the networks with different types of close connected triangle motifs.

network. The $F1$ score improvement ratio of MADNI over the baseline method with varying $\rho$ is shown in Fig. 2(b). Here the improvement ratio is defined as $(F1m - F1_b)/F1_b$, where $F1_m$ is the $F1$ score of MADNI and $F1_b$ is the $F1$ score of the baseline method. It can be observed from the figure that the proposed framework outperforms the baseline even if the target network is close to a random network ($\rho = 0.9$). When the target network becomes more structured, i.e., the ratio of motif-dense network increases, the improvement ratio becomes more significant.

**Adaptivity over various motif patterns**. In order to show that the performance improvement is independent of the specific motif, we examine the performance of our framework on inferring the networks with different types of frequently occurred motifs. Specifically, we consider all the seven close connected triangle motifs in this experiment. We generate the cascades using the Exponential and Rayleigh cascade models [19] and set $\rho = 0.5$. The comparison results are shown in Fig. 3. MADNI consistently performs better than the baseline over networks with different types of frequently occurred motifs, which demonstrates the adaptivity of the proposed framework over various motif patterns.

### 4.2 Experiment on a Real-World Network

In this subsection, we evaluate the proposed framework on a real-world network, i.e., an email communication network of an European Research Institute consisting of 320 nodes and 3031 edges [12]. Similar to the synthetic experiments in Sect. 4.1, we generate $C$ (= 1000, 4000, 10000) cascades on the network as the observations.

In this experiment, we compare the proposed framework with seven methods: NETINF [5], NETINF with community structure prior, NETINF with scale-free prior, NETRATE [19], NETRATE with community structure prior, NETRATE with scale-free prior, and CENI [9]. Here NETINF and NETRATE are classical

**Table 1.** Performance comparison (in terms of $F1$ score) between the proposed MADNI-I/MADNI-R methods and seven competing diffusion network inference algorithms in a real-world network experiment. The proposed methods perform the best.

| Methods | Number of cascades | | |
|---|---|---|---|
| | $C = 1000$ | $C = 4000$ | $C = 10000$ |
| NETINF | 0.5675 | 0.8040 | 0.8333 |
| NETINF + Community structure | 0.5944 | 0.8041 | 0.8363 |
| NETINF + Scale-free | 0.6121 | 0.8044 | 0.8397 |
| NETRATE | 0.6636 | 0.7900 | 0.8350 |
| NETRATE + Community structure | 0.6385 | 0.7900 | 0.8351 |
| NETRATE + Scale-free | 0.6426 | 0.7901 | 0.8431 |
| CENI | 0.3390 | 0.8058 | 0.8517 |
| MADNI-I | 0.6287 | **0.8188** | 0.8464 |
| MADNI-R | **0.6685** | 0.7998 | **0.8600** |

network inference methods, community and scale-free structure are representative structural priors, and CENI is a state-of-the-art network inference algorithm. For the proposed framework, NETINF and NETRATE are employed to learn the diffusion network, i.e., maximize $\big(L(O|A) - R(Z)\big)$, respectively. Therefore, in this experiment, we name the methods generated from the proposed framework as MADNI-I (corresponding to NETINF) and MADNI-R (corresponding to NETRATE), respectively.

Table 1 lists the performance (in terms of $F1$ score) of the proposed MADNI-I/MADNI-R methods and the aforementioned seven competing algorithms under different numbers of cascades. The proposed methods achieve the best performance under all settings of $C$, indicating that the motif prior is powerful in characterizing the complex structure of real-world networks.

### 4.3 Experiment on Real-World Cascades

In this subsection, we evaluate the performance of the proposed framework on a real-world information cascade dataset, i.e., MemeTracker dataset [11]. MemeTracker collects the quotes and phrases posted by the mass medium and Blog sites. This dataset contains 1.5 million news articles and Blog from August 2008 to May 2009. The articles may include hyperlinks of their sources and thus the information propagation can be tracked by the flow of hyperlinks. A site publishes a piece of information with corresponding hyperlink. Sites receives this piece of information would publish similar information and link to their sources. Thus a collection of hyperlinks with time stamps could be regarded as a hyperlink cascade. We construct the hyperlink cascades from top 500 mass media and Blog sites. The total number of cascades is 11262. We test the performance of the proposed methods as well as NETINF, NETINF+Community

**Table 2.** Performance comparison (in terms of $F1$ score) between the proposed MADNI-I/MADNI-R methods and five competing diffusion network inference algorithms in a real-world cascade experiment. The proposed MADNI-I performs the best.

| Methods | Number of cascades | |
|---|---|---|
| | Sub (C = 4000 ) | All (C = 11262) |
| NETINF | 0.2414 | 0.3879 |
| NETINF + Community structure | 0.2425 | 0.3460 |
| NETINF + Scale-free | 0.2730 | 0.3800 |
| NETRATE | 0.2455 | 0.2608 |
| CENI | 0.2538 | 0.2873 |
| MADNI-I | **0.2746** | **0.3885** |
| MADNI-R | 0.2472 | 0.2959 |

structure, NETINF+Scale-free, NETRATE, and CENI with 4000 cascades and all the 11262 cascades, respectively.

Table 2 shows the $F1$ scores of the proposed methods and five competing algorithms under different numbers of cascades. By modeling the motif-prior in the network inference procedure, the proposed MADNI-I performs better than the other competing algorithms under both settings of $C$.

## 5 Conclusion

In this paper, we presented a novel MADNI framework, which mines the motif patterns of the underlying diffusion network and incorporates the uncovered motifs into the network inference procedure via a reweighted motif regularization. By taking the network motifs into consideration, the proposed framework achieves the best performance on both synthetic and real-world datasets.

Future work will be explored from two aspects. First, in the current work, we have only considered the closed triangle motifs in the network inference as they are elementary. In order to better characterize the network structure, more complex motifs such as the higher-order ones should also be taken into account. Second, we will further validate the generalization ability and the flexibility of the proposed framework by incorporating motifs into different baseline methods and comparing the proposed framework to more network inference approaches with various kinds of structural priors.

**Acknowledgment.** The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions. This work was supported in part by the National Natural Science Foundation of China under Grant 61503317, in part by the grants from the Research Grant Council of Hong Kong SAR under Projects RGC/HKBU12202415 and RGC/HKBU12202417, in part by the the Science and Technology Research and Development Fund of Shenzhen with Project Code

JCYJ20170307161544087, and in part by the Faculty Research Grant of Hong Kong Baptist University (HKBU) under Project FRG2/16-17/032.

## References


1. Alon, U.: Network motifs: theory and experimental approaches. Nat. Rev. Genet. **8**(6), 450–461 (2007)
2. Benson, A.R., Gleich, D.F., Leskovec, J.: Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
3. Farajtabar, M., Rodriguez, M.G., Zamani, M., Du, N., Zha, H., Song, L.: Back to the past: source identification in diffusion networks from partially observed cascades. In: Proceedings of 18th AISTATS, pp. 232–240 (2015)
4. Goentoro, L., Shoval, O., Kirschner, M.W., Alon, U.: The incoherent feedforward loop can provide fold-change detection in gene regulation. Mol. Cell **36**(5), 894–899 (2009)
5. Gomez Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of 16th SIGKDD, pp. 1019–1028 (2010)
6. Gomez Rodriguez, M., Leskovec, J., Schölkopf, B.: Structure and dynamics of information pathways in online media. In: Proceedings of 6th ACM WSDM, pp. 23–32 (2013)
7. He, X., Liu, Y.: Not enough data?: Joint inferring multiple diffusion networks via network generation priors. In: Proceedings of 10th WSDM, pp. 465–474 (2017)
8. Hosseini, S.M.J., Lee, S.I.: Learning sparse gaussian graphical models with overlapping blocks. In: NIPS, vol. 30, pp. 3801–3809 (2016)
9. Hu, Q., Xie, S., Lin, S., Wang, S., Philip, S.Y.: Clustering embedded approaches for efficient information network inference. Data Sci. Eng. **1**(1), 29–40 (2016)
10. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web **1**(1), 5 (2007)
11. Leskovec, J., Backstrom, L., Kleinberg, J.: Meme-tracking and the dynamics of the news cycle. In: Proceedings of 15th SIGKDD, pp. 497–506 (2009)
12. Leskovec, J., Krevl, A.: SNAP Datasets: Stanford large network dataset collection, June 2014. http://snap.stanford.edu/data
13. Liu, H., Ioannidis, S., Bhagat, S., Chuah, C.N.: Adding structure: social network inference with graph priors. In: 12th ACM International Workshop on Mining and Learning with Graphs (2016)
14. Liu, K., Cheung, W.K., Liu, J.: Detecting multiple stochastic network motifs in network data. Knowl. Inf. Syst. **42**(1), 49–74 (2015)
15. Liu, Q., Ihler, A.: Learning scale free networks by reweighted l1 regularization. In: Proceedings of 14th AISTATS, pp. 40–48 (2011)
16. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U.: Superfamilies of evolved and designed networks. Science **303**(5663), 1538–1542 (2004)
17. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science **298**(5594), 824–827 (2002)
18. Mukherjee, S., Speed, T.P.: Network inference using informative priors. PNAS **105**(38), 14313–14318 (2008)

19. Rodriguez, M.G., Balduzzi, D., Schölkopf, B.: Uncovering the temporal dynamics of diffusion networks. In: Proceedings of 28th ICML, pp. 561–568 (2011)
20. Rong, Y., Zhu, Q., Cheng, H.: A model-free approach to infer the diffusion network from event cascade. In: Proceedings of 25th CIKM, pp. 1653–1662 (2016)
21. Tang, Q., Sun, S., Xu, J.: Learning scale-free networks by dynamic node specific degree prior. In: Proceedings of 32nd ICML, pp. 2247–2255 (2015)