Analytical Models for Motifs in Temporal Networks

Alexandra Porter
Stanford University
amporter@stanford.edu

Baharan Mirzasoleiman
University of California, Los Angeles
baharan@cs.ucla.edu

Jure Leskovec
Stanford University
jure@cs.stanford.edu

ABSTRACT

Dynamic evolving networks capture temporal relations in domains such as social networks, communication networks, and financial transaction networks. In such networks, temporal motifs, which are repeated sequences of time-stamped edges/transactions, offer valuable information about the networks' evolution and function. However, calculating temporal motif frequencies is computationally expensive as it requires: First, identifying all instances of the static motifs in the static graph induced by the temporal graph. And second, counting the number of subsequences of temporal edges that correspond to a temporal motif and occur within a time window. Since the number of temporal motifs changes over time, finding interesting temporal patterns involves iterative application of the above process over many consecutive time windows. This makes it impractical to scale to large real temporal networks. Here, we develop a fast and accurate model-based method for counting motifs in temporal networks. We first develop the *Temporal Activity State Block Model (TASBM)*, to model temporal motifs in temporal graphs. Then we derive closed-form analytical expressions that allow us to quickly calculate expected motif frequencies and their variances in a given temporal network. Finally, we develop an efficient model fitting method, so that for a given network, we quickly fit the TASMB model and compute motif frequencies. We apply our approach to two real-world networks: a network of financial transactions and an email network. Experiments show that our TASMB framework (1) accurately counts temporal motifs in temporal networks; (2) easily scales to networks with tens of millions of edges/transactions; (3) is about 50x faster than explicit motif counting methods on networks of about 5 million temporal edges, a factor which increases with network size.

CCS CONCEPTS

• **Mathematics of computing** → **Probability and statistics**; **Graph algorithms**; • **Networks** → *Network algorithms*;

KEYWORDS

dynamic networks, temporal motifs

ACM Reference Format:
Alexandra Porter, Baharan Mirzasoleiman, and Jure Leskovec. 2022. Analytical Models for Motifs in Temporal Networks. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3487553.3524669

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04.
https://doi.org/10.1145/3487553.3524669

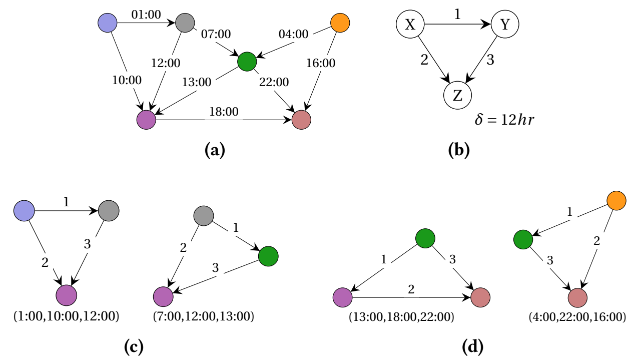


Figure 1: (a) A temporal network with edges appearing over a day, (b) a motif M with a temporal window of 12 hours, (c) examples of M in the network, and d) triangles in the network which are not δ-instances of M, either due to edge order or time between first and last edge.

1 INTRODUCTION

Networks are ubiquitous models for real world systems, with applications ranging from social interactions to protein relationships [15]. Many such systems are not static, but the edges are active only at certain points in time. The networks in which *temporal edges* appear and disappear over time are called time-varying or *temporal* networks. Examples of temporal networks include communication and transaction networks where each link is relatively short or instantaneous, such as phone calls or financial exchanges.

Time dependent and temporal properties can be analyzed on time-varying networks. Extracting recurring and persistent patterns of interaction in temporal networks is of particular interest, as it provides higher order information about the network transformation and functionality [3]. For example, an abundance of triangles in financial transaction networks is associated with anomalies and is identified as the signature of financial crisis [19].

Repeated patterns of interconnections between nodes occurring at a significantly higher frequency than those in randomized networks are called *motifs* [14]. Formally, a *temporal motif* is a subgraph and an ordering on the temporal occurrences of its edges. δ-instances of a temporal motif are instances in which all the temporal edges appear according to the ordering specified by the temporal motif within a time window of length δ [12, 16]. Figure 1 illustrates a small temporal network with δ-instances of a temporal motif M.

Despite the importance of temporal motifs, calculating motif frequencies is computationally expensive and does not scale to large real temporal networks. The reason is that it requires the following two steps. First, identifying all instances of the static motifs in the static graph induced by the temporal graph. Second, counting the number of subsequences of temporal edges that correspond to a temporal motif and occur within a time window of specific length. Both these steps are computationally non-trivial and hard to implement. In addition, as the number of temporal motifs changes over time, finding interesting temporal patterns involves iterative application of the above process over many consecutive time windows. This makes it impractical to scale to large real temporal networks.

Here we develop a fast and accurate model-based method for counting temporal motifs in temporal networks. The heart of our approach is that we develop a network model, which is fast to fit to the data and allows us to obtain motif counts in constant time. In particular, we first develop the *Temporal Activity State Block Model (TASBM)*. We then provide an efficient model parameter fitting technique that scales *linearly* with the number of temporal edges in the graph. And last, we develop a *constant time* analytical method for computing expected motif frequencies and their variances in the TASBM model.

We conduct experiments on both synthetic and real-world temporal networks. Results on synthetic networks demonstrate that our approach can accurately calculate the frequencies of temporal motifs over time. At the same time, our method is about 50x faster than explicit motif counting methods [16] on networks of about 5 million temporal edges. In our real-world experiments, we apply our analytical model to a subset of a financial transaction network with 118.7 thousand nodes and 2.9 million temporal edges. In addition, we apply our framework to a subset of an email network with 997 nodes and 307,869 edges, and in the extended version of the paper we apply our framework to a phone call network with 1.2 million nodes and 21.9 million temporal edges. We show that we can identify interesting temporal patterns based on the expected motif counts calculated by our analytical framework.

2 RELATED WORK

In this section, we briefly summarize the related work to dynamic network models, counting temporal motifs, and estimating their expected frequencies in temporal networks.

There is a body of work on modeling dynamic networks, mostly by extending a static model to the dynamic setting [1, 4, 8, 20–22]. Among the existing methods, temporal extensions of stochastic block models (SBMs) are the most relevant to our work. Our work here differs in that we propose a simple model for temporal motifs that allows for linear time fitting, and hence can be maintained online.

Paranjape et. al. [16] proposed a framework for efficiently enumerating and counting the exact number of relatively small temporal motifs which we use as a baseline in our experiments. Other related methods include [5, 7, 12, 18] and those surveyed by [13]. This line of work aims to accurately count or approximate the actual numbers of motif instances in a graph. However, counting temporal motifs is expensive and does not scale to large real temporal networks. In contrast, our work provides an efficient way to analytically compute the expected motif frequencies and their variance in constant time, which leads to significant speedups.

Currently no analytical model of motifs in temporal graphs exists. However, there have been several heuristic/empirical techniques proposed, based on generating a large ensemble of randomized temporal networks [2, 6, 9, 11]. Such approaches are extremely computationally expensive because they requires generating and materializing new networks and then performing motif counting on each of them. In contrast, we develop the first analytical framework for temporal motifs, which does not require expensive shuffling/simulation of network ensembles, and scales well to large temporal networks.

3 NETWORK ACTIVITY MODEL

In this section we describe our Temporal Activity State Block Model (TASBM). Our goal is to construct a model that captures different activity levels of groups of nodes in a temporal network and can be updated efficiently and thus maintained online. As a result, it can be used to describe the network before the full edge set is known. We note that TASBM only aims to capture activity levels of nodes in a temporal network and not the underlying network structure. Hence, it can be applied to various types of networks, including random, power-law, scale-free, and other network models.

We first provide formal definitions of temporal graphs and temporal edges. Formally, a *temporal graph* $G = (V, E)$ can be viewed as a sequence of static directed graphs over the same (static) set V of $n = |V|$ nodes and the set E of $m = |E|$ *temporal edges*. Each *temporal edge* is a timestamped ordered pair of nodes $(e_i = (u, v), t_i), i \in [m]$, where $u, v \in V$ and $t_i \in \mathbb{R}$ is the timestamp at which the edge arrives. For example, in a phone call network, each temporal edge includes the caller, u, the receiver, v, and the time, t_i, at which the call was placed. Multiple temporal edges between the same pair of nodes u and v with different timestamps can exist. We assume that the timestamps t_i are unique so that the temporal edges may be strictly ordered. However, our methods do not rely on this assumption and can easily be adapted to the case where timestamps are not unique, e.g. by considering each possible ordering of edges with a shared timestamp or selecting an order at random.

3.1 Temporal Activity State Block Model

We now propose our Temporal Activity State Block Model (TASBM), which is a temporal variant of the stochastic block model. The *stochastic block model* (SBM) [4] on static networks is defined as dividing the nodes into communities, or blocks, such that a higher proportion of the possible edges within a block occur, compared to those between blocks. TASBM partitions the nodes of the network based on their activity levels, which we define as the rates at which in- and out- edges arrive to nodes. This means that a particular activity state consists of a set of vertices with *similar temporal activity*, meaning nodes within each activity state all have similar rates of out- and in- edge arrivals.

Specifically, TASBM considers partitioning the nodes into: (1) $R = \{1, \cdots, C^{in}\}$, a partitioning of nodes based on their activity level for receiving in-links; and (2) $S = \{1, \cdots, C^{out}\}$, a partitioning of nodes based on their activity level for sending out-links. Within the i^{th} time interval of length T, i.e., for all $t \in [iT, (i+1)T)$, every node u in the network belongs to (1) a partition $r_u(t) \in R$ based on its activity level for receiving in-links in the i^{th} time interval;

and (2) a partition $s_u(t) \in S$ based on its activity level for sending out-links in the i^{th} time interval. Nodes in the same partition have similar rate of sending or receiving temporal edges.

For every $t \in [iT, (i+1)T)$, we model partition assignments $r_u(t) \in R$, and $s_u(t) \in S$ for node $u \in V$ as independent draws from Multinomial distributions parameterized by $\boldsymbol{\pi_R}(t) \in \mathbb{R}^{C^{in}}$ and $\boldsymbol{\pi_S}(t) \in \mathbb{R}^{C^{out}}$.

In addition, we consider a matrix $\boldsymbol{\theta}(t) \in \mathbb{R}^{C^{out} \times C^{in}}$ such that θ_{sr} denotes the rate of temporal edges from nodes in partition $s \in S$ to the nodes in partition $r \in R$ for all $t \in [iT, (i+1)T)$. We model the temporal edges between every pair (u, v) of nodes as independent Poisson draws from Poisson($\theta_{s_u r_v}(t)$), where $\theta_{s_u r_v}(t) = s_u(t) \cdot r_v(t)$. The choice of Poission process enable TASBM to model the arrival of temporal edges with various rates in different time intervals. Formally, every temporal edge $(e_i = (u, v), t_i)$ from the node u in out-link partition $s_u(t)$ to the node v in in-link partition $r_v(t)$ is an independent Poisson draw. I.e., $e_i = (u, v) \sim \text{Poisson}(\theta_{s_u r_v}(t))$. Each activity state in TASBM consists of nodes that are in the same out-link and in-link partitions. This results in $C = C^{out} \times C^{in}$ activity states.

As nodes change their activity level over time, we assign individual nodes to different activity states. Similarly, we model arrivals of temporal edges between each pair of nodes as a Poisson process with a constant parameter on every time interval of length T. Note that the rates of the Poisson processes can vary for each time window, and hence TASBM is able to robustly and efficiently model the bursty arrival of temporal edges that is observed in real temporal networks [10].

3.2 Parameter Inference in TASBM

For all pairs of vertices (u, v), the Poisson process modeling temporal edges between u and v will be parameterized by a constant $\theta_{s_u r_v}(t)$ for all $t \in [iT, (i+1)T)$. For every time interval, we calculate the posterior model parameters $\boldsymbol{\pi_r}(t), \boldsymbol{\pi_s}(t)$ and $\boldsymbol{\theta}(t)$ as:

$$\boldsymbol{\pi_R}(t) = \frac{\boldsymbol{n_R}(t)}{n}, \quad \boldsymbol{\pi_S}(t) = \frac{\boldsymbol{n_S}(t)}{n}, \quad \theta_{sr}(t) = \frac{m_{sr}(t)}{n_s(t) \cdot n_r(t)},$$

where $\boldsymbol{n_R}(t) \in \mathbb{R}^{C^{in}}$ is the number of nodes in different partitions of R, $\boldsymbol{n_S}(t) \in \mathbb{R}^{C^{out}}$ is the number of nodes in different partitions of S, and $m_{sr}(t)$ is the number of temporal edges from nodes in partition $s \in S$ to nodes in partition $r \in R$, for $t \in [iT, (i+1)T)$. Intuitively, the probability of a node belonging to a partition is equal to the fraction of nodes in that partition for $t \in [iT, (i+1)T)$. Furthermore, the probability of an edge occurring between nodes in partition $s \in S$ and partition $r \in R$ at $t \in [iT, (i+1)T)$ is equal to the number of temporal edges between nodes in s and r, over the number of all possible edges between nodes in s and r, within the same interval. Model inference can be done in at most two passes over the edges and in practice, can be well-approximated in one pass. Thus, this method is extremely scalable and runs in time *linear* in the number of temporal edges in the graph.

4 ANALYTICAL MODEL FOR TEMPORAL MOTIFS

Having defined the model, our next task is to "count" the expected number of temporal motifs. In fact, we do not want to count them as counting would mean materializing/sampling many networks from the model and then running expensive motif counting algorithms. Rather, we will analytically derive closed-form expressions that allow us to quickly calculate the expected values and variance of the motif counts.

We first provide formal definitions of temporal motifs and δ-instances of temporal motifs. We then summarize our closed-form solutions to calculate the expectation and variance of the number of motif instances. Finally, we provide the computational complexity of our method and show that it can easily scale to large real-world temporal networks with millions of temporal edges.

4.1 Temporal Network Motifs

Formally, a temporal motif M defines a particular sequence of interactions between a set of nodes over time.

Definition 1 (temporal motif). *A k-node z-edge temporal motif $M = (G_M, \prec_M)$ consists of a graph $G_M = (V_M, E_M)$, such that $|V_M| = k$ and $|E_M| = z$, and a strict total ordering $\prec_M$ on the edges E_M. We index $E_M = \{e'_1, \cdots, e'_z\}$, such that $e'_1 \prec_M e'_2 \prec_M \cdots \prec_M e'_z$.*

Note that multiple interactions between the same pair of nodes may occur in the sequence defined by M, but each edge is indexed and ordered uniquely. In a dynamic network, any subgraph of a temporal graph is a δ-instance of a temporal motif M if 1) it is isomorphic to G_M, 2) the set of its temporal edges follows the same ordering imposed by M, and 3) it occurs within a time window of δ.

Definition 2 (δ-Instance of a temporal motif). *A temporal subgraph $G_s = (V_s, E_s)$, with a set of temporal edges $E_s = \{(e_1, t_1), \cdots, (e_z, t_z)\}$ is a δ-instance of a temporal motif M if 1) isomorphism: there exist an edge-preserving bijection $f : V_s \to V_M$ between nodes of the subgraph and nodes of the motif such that $\forall_{e=(u,v) \in E_s} (f(u), f(v)) \in E_M$; 2) temporal ordering: the edges of the temporal motif occur according to the ordering $\prec_M$, i.e., for the ordered sequence $f(e_h) \prec_M f(e_i) \prec_M \cdots \prec_M f(e_j)$ we get a set of strictly increasing timestamps $t_h < t_i < \cdots < t_j$; and 3) temporal window: all the edges in E_s occur within δ time, i.e. $t_j - t_h \leq \delta$.*

Here, our goal is to derive the expected number (and the variance) of δ-instances of a given temporal motif in a time-varying network.

4.2 Expected Motif Frequencies

Next we provide a closed form solution for the expected δ-instances of temporal motifs in TASBM networks. Derivation details can be found in the full version of the paper [17]. To calculate the expected number of δ-instances of a temporal motif over a time window of length T, we need to calculate the expected number of subgraphs $G_s = (V_s, E_s)$ satisfying the conditions specified in Definition 2. The following Theorem summarizes our main theoretical results.

Theorem 1. *The expected number of δ-instances of a k-node z-edge temporal motif M in a network with a set V of nodes modeled by a TASBM with C states during a time interval $[t_0, t_0 + T)$ for $T \leq \delta$ is*

$$\mathbb{E}[N_M | T \leq \delta] = \mathbb{E}[N_{S^{k,z}_{V,C}}] \cdot Pr(t_0 \leq t_1 < t_2 < \ldots < t_z < t_0 + T),$$

where the first term is the expected number of k-node z-edge isomorphic subgraphs to the motif graph G_M in a temporal network with V nodes modeled by a TASBM with C states; and the second term is the *probability that timestamps of the edges in each isomorphic subgraph occur in the order $\prec_M$ specified by the motif in a time window of length δ.*

We next summarize how the two terms can be computed. Let $\mathcal{A}_{C,k}$ denote the set of activity state assignments for a set of k nodes in TASBM, so $|\mathcal{A}_{C,k}| \leq C^k$. Let n^c be the number of node in activity state $c \in [C]$. For some activity state assignment $A \in \mathcal{A}_{C,k}$, let n^c_A be the number of nodes in A assigned to activity state c.

Let $P(n^c, n^c_A)$ be the number of n^c_A-permutations of n^c, so $P(n^c, n^c_A) = \binom{n^c}{n^c_A} n^c_A!$. Note that $\prod_{c \in [C]} P(n^c, n^c_A)$ is constant for every activity state assignment $A \in \mathcal{A}_{C,k}$.

Lemma 1. *The expected number of k-node subgraphs isomorphic to the motif subgraph G_M (with edges E_M) in a Temporal Activity State Block Model (TASBM) during a time interval $[t_0, t_0 + T)$ for $T \leq \delta$ is*

$$\mathbb{E}[N_{S^{k,z}_{V,C}} | t \in [t_0, t_0 + \delta)] =$$

$$\sum_{A \in \mathcal{A}_{C,k}} \prod_{c \in [C]} P(n^c, n^c_s) \prod_{\substack{u,v \in V_s | R(V_s) = A, \\ (f(u), f(v)) \in E_M}} \int_{t_0}^{t_0+T} \theta_{s_u r_v}(t) dt,$$

where $R(V_s) = A$ is the set of all k-node subgraphs consistent with activity state assignment A.

Next, we consider how to compute the second term in theorem 1. For ease of notation, we subsequently assume that for each subgraph G_S and corresponding bijection $f : V_s \rightarrow V_M$, we have $f(e_1) \prec_M f(e_2) \prec_M \cdots \prec_M f(e_z)$. Therefore, we need to calculate the probability that $t_0 \leq t_1 < t_2 < \cdots < t_z < t_0 + \delta$.

The marginal probability for a temporal edge $e = (u, v)$ from activity state s_u to activity state r_v to occur at time t in the time window $[t_0, t_0 + T)$ is

$$\Theta^{[t_0, t_0+T)}_{e=(u,v)}(t) = \frac{\theta_{s_u r_v}(t)}{\int_{t_0}^{t_0+T} \theta_{s_u r_v}(t') dt'}.$$

Lemma 2. *The probability that temporal edges of a subgraph $G(V_s, E_s)$ occur in the order $\prec_M$ specified by motif M in an interval $[t_0, t_0 + T)$ is*

$$Pr(t_0 \leq t_1 < t_2 < ... < t_z < t_0 + T) =$$

$$\int_{t_0}^{t_0+T} \Theta^{[t_0,t_0+T)}_{e_1}(t_1) \int_{t_1}^{t_0+T} \Theta^{[t_0,t_0+T)}_{e_2}(t_2) \cdots \int_{t_{z-1}}^{t_0+T} \Theta^{[t_0,t_0+T)}_{e_z}(t_z) dt_z dt_{z-1} dt_{z-2} ... dt_1.$$

We next extend our result to the scenario where $T > \delta$. Here, the δ-instances of a temporal motif may have at least one edge occurring in $[t_0, t_0 + T - \delta)$ or they may have all edges occurring in $[t_0 + T - \delta, t_0 + T)$. Hence, to calculate the expected number of instances in T, we take the sum over these cases:

Theorem 2. *The expected number of δ-instances of a temporal motif M in a Temporal Activity State Block Model (TASBM) during a time interval $[t_0, t_0 + T)$ for $T > \delta$ is*

$$\mathbb{E}[N_M | T > \delta] = \mathbb{E}[N_M | T = \delta] + \mathbb{E}[N_M | T > \delta, t_1 < t_0 + T - \delta],$$

where t_0 and t_1 are the timestamps of the first and second edges of the motif instance.

4.3 Variance of Motif Counts

We next discuss how we can use our framework for deriving the variance of the number of motif instances $\mathbb{V}[N_M]$. I.e., $\mathbb{V}[N_M] = \mathbb{E}[N_M^2] - \mathbb{E}[N_M]^2$. While we can simply calculate $\mathbb{E}[N_M]^2$ using Theorems 1 and 2, computing $\mathbb{E}[N_M^2]$ involves calculating the expected number of *pairs* of δ-instances of motif M. A pair of instances can overlap in up to k vertices and up to z temporal edges.

Let S_1, S_2 be a pair of of δ-instances of M. The time interval that the edges $E_{S_1} \cup E_{S_2}$ of both instances may occur is within an interval $[t_0 + \delta, t_0 + 2\delta)$, depending on which temporal edges, if any, are shared. We denote by $\mathbb{E}_\delta$ and $\mathbb{E}_{\delta'}$ expected δ- and δ'-instances of motif M, respectively. Then by linearity of expectation, we get

$$\mathbb{E}[N_M^2] = \sum_{\substack{(S_1, S_2): \\ E_{S_1} \cap E_{S_2} = \emptyset}} \mathbb{E}_\delta[N_{S_1} | t \in [t_0, t_0 + T)] \mathbb{E}_\delta[N_{S_2} | t \in [t_0, t_0 + T)] + \sum_{\substack{(S_1, S_2): \\ E_{S_1} \cap E_{S_2} \neq \emptyset}} \sum_{\delta' \in [t_0 + \delta, t_0 + 2\delta)} \mathbb{E}_{\delta'}[N_{S_1 \cup S_2} | t \in [t_0, t_0 + T)].$$

4.4 Computational Complexity

The computational complexity of our method to compute the expected number of δ-instances of a k-node z-edge temporal motif in a TASBM with C activity states is $O(C^k)$ where C is the number of blocks (usually < 10) and k is motif size (usually < 10), and assuming the cost of computing each integral is $O(1)$. Notice the motif size k is constant relative to the size of the network and in most cases the number of blocks C can be as well, so we get a computational complexity of motif counting to be $O(1)$. Due to its low computational complexity, our analytical method can easily scale to large real-world temporal networks with millions of temporal edges.

5 EXPERIMENTS

In this section, we present the results of applying our framework [1] to calculate the expected motif frequencies in synthetic and real-world temporal networks. We compare the expected number of motif instances to the number of observed motif instances counted by the method of [16], which is the most efficient method for counting the exact number of (relatively small) temporal motifs. We focus on expected counts for motifs with 2 or 3 nodes and 3 edges (Figure 2), and note that our analytical framework scales well to calculate the expected frequencies for larger motifs.

5.1 Accuracy on Synthetic Networks

Our synthetic networks are generated according to the TASBM with $C = C^{out} \times C^{in}$ activity states, for varying number of out-edge rates, C^{out}, and in-edge rates, C^{in}. Each node is assigned to one of the C activity states chosen uniformly at random. We consider a different rate of edge arrivals between any pair of activity states. For each pair of nodes (u, v), temporal edges are then sampled according to a Poisson process with the rate $\theta_{s_u r_v}$.

Accuracy of Model Inference. We first show that we can accurately infer the TASBM parameters used to generate the synthetic

[1] Our C++ implementation of the model can be found at: https://github.com/aporter468/motifsanalyticalmodel

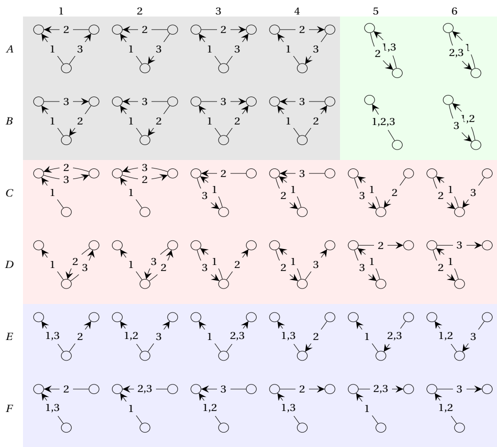


Figure 2: All 2- and 3- node motifs with 3 edges, shaded by main structural feature: triangles (A1-4, B1-4; grey), Two-Node (A5,6 B5,6; green), Reciprocated edge (C1-6,D1-6; red), and Double edge (E1-6,F1-6; blue).

Table 1: MSRE for varying number of groups C used by TASBM using $T=10K$, $\delta=5K$. The values are calculated over 30 networks generated with 300 nodes and $C^{out}=C^{in}=\sqrt{C}$.

C	Triangles	Two Vertex	Reciprocated	Double Edge
1	0.229	0.381	0.147	0.381
4	1.99e-05	4.35e-05	2.84e-05	1.69e-05
9	1.89e-05	4.26e-05	2.78e-05	1.60e-05
16	1.04e-05	3.59e-05	2.25e-05	7.90e-06
25	1.04e-05	3.59e-05	2.25e-05	7.91e-06

network and thus accurately use our analytical approach to determine expected number of motifs. We measure accuracy of our model over a set of r synthetic networks, using Mean Squared Relative Error: $MSRE = \frac{1}{r}\sum_{i=1}^{r}\left(\frac{N_M^i - N_M}{N_M^i}\right)^2$ where N_M^i is the actual number of motif instances counted using the method of [16] in the i-th generated network, and N_M is the expected motif frequency calculated by our framework.

Table 1 shows MSRE for 30 networks with 300 nodes generated using TASBM with $C^{out}=C^{in}$. For generating out-edges, we divide the nodes to partitions of size $|S|$ =10, 30, 60, 80, and 120, with out-rates of 1e-7, 1e-6, 1e-5, 1e-4, and 1e-3. For generating in-edges, we have all the nodes in one partition, i.e. $|R| = 1$, hence having in-rate of 11111e-3. Generated networks have an average of 384,580 edges. To infer the parameters, we vary the number of activity states in our framework for calculating the expected motif frequencies from $C^{out}=C^{in}=1$ to $C^{out}=C^{in}=5$ and used $T=10K$ and $\delta=5K$ time units. The error quickly vanishes when the model is allowed to use a higher number of activity states, but the improvements from increasing the number of activity states quickly diminish. Using only $C^{out}=C^{in}=2$ partitions to calculate the expected frequencies, we get almost the same accuracy as using $C^{out}=C^{in}=5$ partitions. This indicates that while real data is likely to have a large variety of node activity levels, a relatively small value of C can be used to calculate accurate expected motif counts.

Robustness of Model to Hyper-Parameter Choices. Next we investigate robustness of our method to choices of hyper-parameters. Figure 3 compares MSRE for 30 networks with 100 nodes generated using TASBM with $C^{out}=C^{in}=3$. Here, for generating out-links we divide the nodes into partitions of sizes 10, 30, and 60 and use the initial out-rates of 5e-6, 1e-4, and 1e-3, respectively. The rates were chosen to be sufficiently distinct and to generate sufficiently large edge volumes without motif counts exceeding the maximum capacity of the motif counter from [16]. Figure 3 shows the accuracy of our model as average degree, motif window δ, and time window T are varied. In all cases, MSRE converges to zero as the varied parameter is increased.

5.2 Run Time Compared to Motif Counting

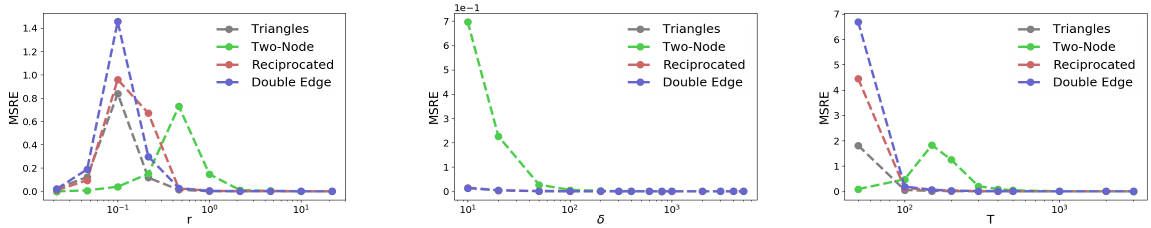
We next show how the run time of our analytical model compares to the motif counting methods in [16] for all 2- and 3- node motifs with 3 edges. We measure the entire execution time of our implementation, including inference of the TASBM parameters and computation of the 36 expected motif frequencies. We compare this to the execution time of motif counting as measured by the implementation in [16][2]. Run times were measured on synthetic networks generated using $C = 36$ for values of T ranging from 50 thousand to 28 million, with corresponding edge counts ranging from 9 thousand to 5 million. As shown in Figure 6, on smaller graphs our analytical model takes slightly longer to execute compared to motif counting. However, the run time of our model hardly grows as networks increase in size, while the computational cost of motif counting grows exponentially with the number of temporal edges. In particular, for larger network of 5 million temporal edges, our method is about 50x faster than explicit motif counting methods.

5.3 Accuracy on Real-World Networks

In our real-world experiments, we apply TASBM to a financial transaction network and an email network (additional experiments on a phone call network can be found in the full verison of the paper). We show that our model matches the trends in motif counts in the real data. Since we are interested in modeling temporal dynamics of the networks, we preprocess the financial transaction and email networks by removing low-degree (< 10% the largest degree) nodes and focusing on the largest connected community.

Financial Transaction Network. In our first real-world experiment, we applied our framework to model the motif counts in a small European country's financial transaction network. The data is collected from the entire country's transaction log for all transactions larger than 50K Euros over 10 years from 2008 to 2018, and includes 118,739 nodes and 2,982,049 temporal edges. As edges do not occur on weekends, we do not count them toward values of T and δ or in computing edge rates. Figure 4 compares the ground-truth motif counts and the values computed by our model with $\delta = T = 90$ weekdays for motifs F1 and A6. Notice that TASBM is able to accurately calculate the motif counts over time. It is notable

[2]Code found here: http://snap.stanford.edu/temporal-motifs/code.html



(a) (b) (c)

Figure 3: MSRE for the expected motif counts over 30 synthetic networks with 100 nodes for varying (a) average degree with $\delta = 5K$, $T = 10K$, (b) motif window δ, with $T = 10K$, and (c) time window T, with $\delta = 5K$. See Figure 2 for motif names.

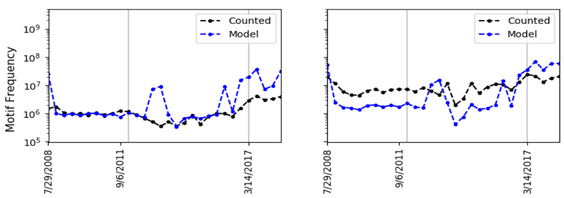


Figure 4: Financial transaction network, $\delta = T = 90$ days. Motifs F1 and A6. Notice the model accurately calculate the motif frequencies over time.

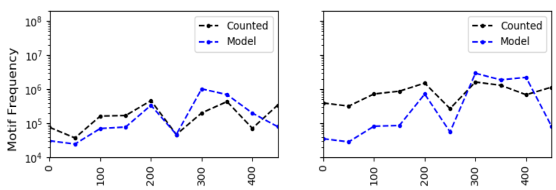


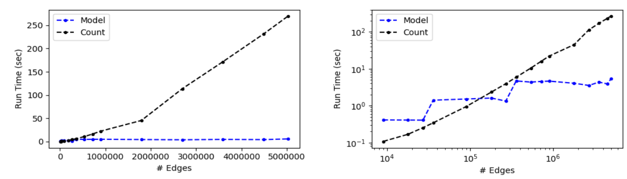
Figure 5: Email network, $\delta = T = 50$ days. Motifs F1 and A6. Notice the model accurately tracks motif counts over time.

that in September of 2011, when a financial crisis hits the country, the number of modeled motifs dropped significantly. It can also be observed that the network starts to recover around March 2017.

Email Network. We next applied our model to a network of emails exchanged within a European research institution. We use the set of 307,869 temporal edges over 977 nodes, which appear in the 500 days beginning in October of 2003. We modeled and counted frequencies of 2– and 3– node motifs with 3 edges at a time scale of $\delta = T = 50$ days with 10 intervals. As shown in Figure 5 our model follows the trends in motif counts on the entire network.

6 DISCUSSION

We have developed a fast and accurate model-based method to determine the expected number as well as the variance of motifs in a temporal network. We developed an efficient parameter inference technique as well as provided closed form solutions for the expected motif frequencies in the general case where temporal edges appear with distinct rates between different pairs of nodes in various activity states, and the arrival rate of temporal edges between every pair of activity states may change over time. We



(a) (b)

Figure 6: Average run time of the TASBM model implementation and motif counting algorithm of [16] on synthetic graphs of increasing size, shown on (a) linear and (b) logarithmic scales. The steps in run time visible in (b) for the TASBM implementation correspond to points at which the value of C is increased.

demonstrated the effectiveness of our Temporal Activity State Block Model combined with our analytical model of temporal motifs for computing expected motif frequencies in temporal networks.

REFERENCES

[1] A. Ahmed and E. P. Xing. Recovering time-varying networks of dependencies in social and biological studies. *PNAS*, 106(29):11878–11883, 2009.
[2] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza. Dynamical patterns of cattle trade movements. *PloS one*, 6(5):e19869, 2011.
[3] A. R. Benson, D. F. Gleich, and J. Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.
[4] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.
[5] V. Dias, C. H. Teixeira, D. Guedes, W. Meira, and S. Parthasarathy. Fractal: A general-purpose graph pattern mining system. In *Proceedings of the 2019 International Conference on Management of Data*, pages 1357–1374, 2019.
[6] T. Donker, J. Wallinga, and H. Grundmann. Dispersal of antibiotic-resistant high-risk clones by hospital networks: changing the patient direction can make all the difference. *Journal of Hospital Infection*, 86(1):34–41, 2014.
[7] S. Gurukar, S. Ranu, and B. Ravindran. Commit: A scalable approach to mining communication motifs from dynamic networks. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 475–489, 2015.
[8] Q. Ho, L. Song, and E. Xing. Evolving cluster mixed-membership blockmodel for time-evolving networks. In *AISTATS*, pages 342–350, 2011.
[9] P. Holme and F. Liljeros. Birth and death of links control disease spreading in empirical contact networks. *Scientific reports*, 4:4999, 2014.
[10] J. Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4):373–397, 2003.
[11] M. Li, V. D. Rao, T. Gernat, and H. Dankowicz. Lifetime-preserving reference models for characterizing spreading dynamics on temporal networks. *Scientific reports*, 8(1):709, 2018.
[12] P. Liu, A. Benson, and M. Charikar. A sampling framework for counting temporal motifs. *arXiv preprint arXiv:1810.00980*, 2018.
[13] P. Liu, V. Guarrasi, and A. E. Sariyuce. Temporal network motifs: Models, limitations, evaluation. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[15] M. E. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[16] A. Paranjape, A. R. Benson, and J. Leskovec. Motifs in temporal networks. In *WSDM*, pages 601–610. ACM, 2017.

[17] A. Porter, B. Mirzasoleiman, and J. Leskovec. Analytical models for motifs in temporal networks: Discovering trends and anomalies. *arXiv preprint arXiv:2112.14871*, 2021.

[18] U. Redmond and P. Cunningham. Temporal subgraph isomorphism. In *ASONAM*, pages 1451–1452. IEEE, 2013.

[19] T. Squartini, I. Van Lelyveld, and D. Garlaschelli. Early-warning signals of topological collapse in interbank networks. *Scientific reports*, 3:3357, 2013.

[20] A. H. Westveld, P. D. Hoff, et al. A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, 5(2A):843–872, 2011.

[21] K. S. Xu and A. O. Hero. Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *SBP*, pages 201–210. Springer, 2013.

[22] T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin. Detecting communities and their evolutions in dynamic social networks-a bayesian approach. *Machine learning*, 82(2):157–189, 2011.