Sharp Thresholds in Random Simple Temporal Graphs

Arnaud Casteigts* Michael Raskin† Malte Renken‡ Viktor Zamaraev§

November 10, 2020

Abstract

A graph whose edges only appear at certain points in time is called a temporal graph (among other names). Such a graph is temporally connected if each ordered pair of vertices is connected by a path which traverses edges in chronological order (i.e., a temporal path). In this paper, we consider a simple model of random temporal graph, obtained by assigning to every edge of an Erdős–Rényi random graph $G_{n,p}$ a uniformly random presence time in the real interval [0, 1].

It turns out that this model exhibits a surprisingly regular sequence of thresholds related to temporal reachability. In particular, we show that at $p = \log n/n$ any fixed pair of vertices can a.a.s. reach each other, at $2 \log n/n$ at least one vertex (and in fact, any fixed node) can a.a.s. reach all others, and at $3 \log n/n$ all the vertices can a.a.s. reach each other (i.e., the graph is temporally connected). All these thresholds are sharp. In addition, at $p = 4 \log n/n$ the graph contains a spanning subgraph of minimum possible size that preserves temporal connectivity, i.e., it admits a temporal spanner with $2n - 4$ edges.

Another contribution of this paper is to connect our model and the above results with existing ones in several other topics, including gossip theory (rumor spreading), population protocols (with sequential random scheduler), and edge-ordered graphs. In particular, our analyses can be extended to strengthen several known results in these fields.

1 Introduction

A temporal graph is a graph whose edges are present only at certain times. These graphs can be modeled in various ways, a classical option being as an edge-labeled graph $\mathcal{G} = (G, \lambda)$, where G is a static graph and λ encodes the presence times of the edges of G. Temporal graphs have been extensively studied in the past two decades, motivated by the modeling of dynamic networks in social network analysis, epidemics, biology, and communication networks. One of the central concepts in these graphs is that of a *temporal path*, which is a path running over a non-decreasing (or increasing) sequence of timestamps. Reachability based on temporal paths poses a number of fundamental questions. In particular, it is non-transitive, with dramatic effects on the most basic notions. For instance, Kempe, Kleinberg, and Kumar [15] showed that deciding whether k vertex-disjoint temporal paths exist between two given vertices of a

*LaBRI, CNRS, University of Bordeaux, Bordeaux INP, France. Email: arnaud.casteigts@labri.fr. ORCID: 0000-0002-7819-7013. This research was supported by ANR project ESTATE (ANR-16-CE25-0009-03).

†Department of Informatics, Technical University of Munich, Germany. Email: raskin@in.tum.de. ORCID:0000-0002-6660-5673. The author has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 787367 (PaVeS).

‡Technical University of Berlin, Germany. Email: m.renken@tu-berlin.de. ORCID: 0000-0002-1450-1901. The author is supported by the German Research Foundation (DFG) under the project MATE (NI 369/17).

§Department of Computer Science, University of Liverpool, UK. Email: Viktor.Zamaraev@liverpool.ac.uk. ORCID: 0000-0001-5755-4141.

temporal graph is NP-hard, the analog problem in static graphs being polynomial-time solvable. Similarly, Bhadra and Ferreira [5] showed that computing a maximum temporal component is as hard as finding a maximum clique in a static graph. Indeed, maximal components in temporal graphs do not form equivalence classes with respect to the reachability relation.

The very definition of some basic concepts may not be easy to formulate. An interesting example is the one of a spanning tree, i.e., a cycle-free subset of edges which connects all the vertices of a graph. In static graphs, not only the existence of a spanning tree is guaranteed, but its computation is also straightforward. In contrast, it has been shown recently [3] that certain temporal graphs do not even admit *sparse* spanners, i.e., spanning subgraphs with $o(n^2)$ many edges that preserve temporal connectivity of the input graph. However, such spanners are guaranteed in particular cases, for example when the input graph is a complete graph [11]. More generally, understanding what conditions enable the existence of a sparse spanner is essentially an open problem taking its roots yet in the seminal work of Kempe, Kleinberg, and Kumar [15].

In this work, we are interested in studying the properties of temporal reachability from a probabilistic point of view. To this goal, we study a simple and natural model of random temporal graphs which intends to play an analogous role to the Erdős–Rényi model for static graphs. A temporal graph $\mathcal{G} = (G, \lambda)$ is called *simple* if λ is single-valued; in other words, every edge of G is present at a single time. Given a number of nodes n and a probability p, a *random simple temporal graph* (RSTG, for short) is the graph $\mathcal{G} = (G, \lambda)$ where G is an Erdős–Rényi graph $G = (V, E) \sim G_{n,p}$ and $\lambda : E \to [0, 1]$ assigns a unique uniformly random presence time from the unit interval independently to every edge of G. Observe that the reachability in this model only depends on the relative order of edge labels. This model could thus be equivalently defined through a random edge-ordering of an Erdős–Rényi graph (the former definition turns out to be more convenient for our analyses).

RSTGs are related to classical processes in gossip theory and population protocols. In a certain sense, they *encode* information pertaining to the ordering of interactions in these models. A significant difference to most of these models is that there are no repeated interactions. Nonetheless, the results we obtain in the RSTG model can be transferred to these models as well, where they provide new results and strengthen some known ones.

1.1 Technical results

Inspired by connectivity questions on Erdős–Rényi graphs, we investigate natural analogs of these questions in RSTGs. The fact that temporal reachability is neither transitive nor symmetrical implies the existence of a number of different thresholds for connectivity. More precisely, we consider the following five gradual properties: (1) a fixed node can reach another fixed node asymptotically almost surely (a.a.s.); (2) a.a.s. at least one node can reach all the others; (3) a fixed node can a.a.s. reach all the others; (4) every node can reach all the others a.a.s. (i.e., the graph is a.a.s. temporally connected); and (5) the graph a.a.s. admits a temporal spanner with $2n - 4$ edges, which is the minimum possible due to a classical combinatorial results (see e.g. [9]). Quite surprisingly, the model exhibits a regular set of incremental thresholds for these properties as follows. The first property occurs at $p = \log n/n$ (where log is the natural logarithm), the second and third occur at $p = 2 \log n/n$, and the fourth at $p = 3 \log n/n$. All these thresholds are sharp; that is, for the threshold factors $c \in \{1, 2, 3\}$, the property a.a.s. does not hold at $p = (c - o(1)) \log n/n$ and holds a.a.s. at $p = (c + o(1)) \log n/n$. These sharp thresholds are summarized in Table 1. As for the existence of an optimal spanner (fifth property), we prove that such a spanner exists at $p = 4 \log n/n$ and conjecture that this is a sharp threshold as well.

1.2 Significance of the results

From the point of view of static graphs, there is no distinction between properties (2), (3), (4), and (5), even in the deterministic case. Indeed, as soon as a node u can reach all other nodes, we obtain by transitivity and symmetry of the reachability relation that every node can reach all others (through u, for example), and a spanning tree of $n-1$ edges (which is a size-optimal connected spanning subgraph) exists unconditionally. All these properties occur at $p = (1 + o(1)) \log n/n$ in the Erdős–Rényi model. Property (1), on the other hand, is known to occur before, but without obeying a sharp threshold [13, Section 2.2]. In view of this context, the fact that the temporal analogs of the first four properties occur at three distinct thresholds and that these thresholds are all sharp is quite significant. These thresholds can be seen as a fine-grained measure of the discrepancy between static and temporal connectivity.

Due to their simplicity, RSTGs are natural temporal analogs to Erdős–Rényi graphs. Despite the strong limitations that a single time label per edge imposes on the resulting graph, it turns out that RSTGs also capture classical phenomena observed in the literature on gossip theory and population protocols. A classical question in gossip theory asks how long it takes for a rumor to spread among a set of agents through random phone calls. If the calls are chosen uniformly at random without repetition, then the ordering of the calls defines a process that is exactly captured by RSTGs. However, most authors study calls with repetition. It turns out that, in this case, the ordering of phone calls defines a temporal graph whose early evolution is sufficiently close to RSTGs for the reachability thresholds to remain valid (in a sense that will be made precise). A similar observation applies in the more recent field of population protocols (see e.g. [2]), where a set of agents carries out a distributed task through pairwise interactions, these interactions being typically chosen by a central scheduler. Unless the scheduler is adversarial, it is generally considered that the interactions are chosen uniformly at random with repetition, which corresponds exactly to the above gossip setting. In the following, we refer to this setting as the randomized `ANY` model (following recent notations from [12]). Perhaps surprisingly, few connections have been made between gossip theory and population protocols, and it is not uncommon to see existing gossip results being re-discovered independently in the analysis of population protocols. Let us briefly review how our results connect to state-of-the-art analyses in the randomized `ANY` model, irrespective of the field in which they were carried out.

In a sequence of three papers from the '70s [19, 8, 14], with the common title "*Random exchanges of information*" (but different authors), the analogs of two of the above temporal graph properties were studied in the randomized `ANY` model. Namely, the state where a fixed agent receives the rumors from all of the other agents (by symmetry, this is "equivalent" to the state when all agents receive a fixed rumor, which corresponds to our third temporal graph property above), and the state where all agents receive the rumor from all agents, which is the analog for temporal connectivity. In both cases, asymptotics for the *expected* number of calls to reach the desired state were established. For the former it is $n \log n$, which indeed corresponds to the expected number of edges in an RSTG with $p = 2 \log n/n$, and the latter is $1.5n \log n$, which corresponds to $p = 3 \log n/n$. It was recently shown that for the former state, the actual number of calls is indeed concentrated around its expected value [18], and for the latter state, the actual number of calls does not exceed its expected value a.a.s. [10][1]

By translating our results for RSTGs to the randomized `ANY` model, we do not only establish concentration results (i.e., sharp thresholds) for the durations for the above two states, but also obtain similar results for other natural states including complete propagation of at least one rumor, and exchange of rumors between two fixed agents. Furthermore, we obtain the same

[1]While Lemma A.7 in [10] claims a weaker result, an intermediate step in the proof directly implies the upper bound of $(1.5 + o(1))n \log n$.

Property	Shorthand	Sharp threshold	Reference
Point-to-point Reachability	$\forall u \forall v$ **a.a.s.** $u \rightsquigarrow v$	$\log n/n$	Theorem 5.1
First Temporal Source	**a.a.s.** $\exists u \forall v \; u \rightsquigarrow v$	$2 \log n/n$	Theorem 5.3
Temporal Source	$\forall u$ **a.a.s.** $\forall v \; u \rightsquigarrow v$	$2 \log n/n$	Theorem 5.2
Temporal Connectivity	**a.a.s.** $\forall u \forall v \; u \rightsquigarrow v$	$3 \log n/n$	Theorem 5.4

Table 1: Sharp thresholds for connectivity properties in random temporal graphs. The notation $u \rightsquigarrow v$ denotes the existence of a temporal path between u and v.

results in a natural rumor propagation model with dependencies known as the randomized "call-once" (`CO`) model. As the name suggests, in this model, every next call is chosen uniformly at random among all calls that did not happen before. The independence between past and future calls in `ANY` model is convenient and heavily utilized in the analysis. In contrast, absence of such independence in the randomized `CO` model requires different approaches for model analysis, and in fact not much is known about the randomized `CO` model according to [12].

Due to the initial purpose, our techniques are originally designed to deal with dependencies and therefore the results for RSTGs directly translate to the randomized `CO` model. Moreover, the flexibility of our techniques allows us to transfer all our results to the randomized `ANY` model with minimal computational changes. Thus, our results show that within the two models information essentially propagates at the same speed.

1.3 Organization of the document

The main definitions are given in Section 2, including basic concepts and notations in temporal graphs. Section 3 presents an analysis that obtains weaker bounds than our actual results using simpler arguments; while proposed as a warm-up, these bounds are utilized in subsequent analysis. Section 4 develops our main tools, by introducing an algorithm that grows a *foremost tree* (i.e., a tree of time-optimal temporal paths) in a given temporal graph and analyzing the execution of this algorithm on a typical RSTG. Section 5 applies the tools from the previous section to obtain the main results, namely the claimed thresholds in RSTGs. Section 6 describes the adaptation of our analyses in models coming from gossip theory and population protocols. We also explain in detail how our results strengthen known results in these fields. Finally, Section 7 concludes with some remarks and open questions.

2 Preliminaries

2.1 Temporal graphs

A *temporal graph* $\mathcal{G}$ is a pair (G, λ), where $G = (V, E)$ is a simple undirected graph and λ is a function that assigns to every edge e of G a finite set of elements from some totally ordered set $\mathbb{T}$. The graph G is called the *underlying graph* of $\mathcal{G}$ and the elements of $\lambda(e)$ are called *time labels of* e. The temporal graph $\mathcal{G}$ is *simple* if every edge of G is assigned exactly one time label, i.e., $|\lambda(e)| = 1$ for every $e \in E$. We will mostly focus on simple temporal graphs. With a slight abuse of notation we write $\lambda(e)$ to denote the unique time label of an edge $e \in E$.

A *temporal* (u, v)*-path* or a *temporal path* from u to v in $\mathcal{G}$ is a path $u = u_0, u_1, \ldots, u_\ell = v$ in G such that there are labels $\lambda_i \in \lambda(u_{i-1}u_i)$ with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\ell$. The label λ_ℓ is then called the *arrival time* of the temporal path. A vertex $v \in V$ is called a *temporal source of* $\mathcal{G}$ if every other vertex can be reached from v by a temporal path. Similarly, vertex v is called a *temporal sink of* $\mathcal{G}$ if every other vertex can reach v by a temporal path. The temporal graph

$\mathcal{G}$ is *temporally connected* if every vertex can reach any other vertex by a temporal path.

A temporal graph $\mathcal{G}' = (G', \lambda')$ is a *temporal subgraph* of $\mathcal{G}$ if G' is a subgraph of G and $\lambda'(e) \subseteq \lambda(e)$ for every edge e of G'. If $\mathcal{G}'$ is temporally connected and if $V(\mathcal{G}') = V(\mathcal{G})$, then it is called a *(temporal) spanner* of $\mathcal{G}$. Another important example of a temporal subgraph is the restriction of $\mathcal{G} = (G, \lambda)$ to a time interval $[a, b]$, which is defined as $\mathcal{G}_{[a,b]} = (G', \lambda')$ where $\lambda'(e) = \lambda(e) \cap [a, b]$ and $G' = (V, \{e \in E \mid \lambda'(e) \neq \emptyset\})$.

2.2 Random simple temporal graphs (RSTGs)

The most basic and commonly studied model of random static graphs is the Erdős–Rényi model denoted $G_{n,p}$, in which there are n vertices and every two of them are connected by an edge independently with probability $p \in [0, 1]$. A possible way of turning $G_{n,p}$ into a model of random *temporal* graphs is by choosing uniformly at random a total order on the set of edges. More specifically, a random temporal graph (G, λ) can be obtained by first drawing the underlying graph G from $G_{n,p}$ and then drawing λ uniformly at random from all bijections $E(G) \to \{1, 2, \ldots, |E(G)|\}$. For technical convenience, however, we work with a slightly different but equivalent model which we denote $\mathcal{F}_{n,p}$. In this model, the underlying graph is drawn from $G_{n,p}$ as before, but the labeling function λ now maps every edge of $E(G)$ to an independent and uniformly distributed label in the real interval $[0, 1]$. Since with probability 1 no two time labels are equal, this induces a total order on the edges, and by symmetry, all orders are equally likely, thus $\mathcal{F}_{n,p}$ is equivalent to the above model of random edge-orderings of $G_{n,p}$ (this equivalence has been used in some recent works on edge-ordered graphs [1, 16]). We refer to such a graph as a random simple temporal graph (RSTG), or just a random temporal graph, which is simple by default.

2.3 Degrees of reachability

A graph property is said to hold asymptotically almost surely (a.a.s.) if the probability that it is satisfied converges to 1 as n goes to infinity. A function $p = p(n)$ is a *sharp threshold* for a temporal graph property $\mathcal{P}$ if for every $\varepsilon > 0$ a random temporal graph from $\mathcal{F}_{n,(1-\varepsilon)p}$ does not have property $\mathcal{P}$ a.a.s., while a random temporal graph from $\mathcal{F}_{n,(1+\varepsilon)p}$ has $\mathcal{P}$ a.a.s.

We study the following fundamental properties related to temporal reachability:

1. **Point-to-point Reachability**. The property that a fixed pair (u, v) of vertices has a temporal path from u to v.

2. **First Temporal Source**. The property that $\mathcal{G}$ contains at least one temporal source.

3. **Temporal Source**. The property that a fixed vertex v is a temporal source, i.e., every vertex in $\mathcal{G}$ can be reached from v by a temporal path.

4. **Temporal Connectivity**. The property that $\mathcal{G}$ is temporally connected, which is equivalent to the property that all vertices in $\mathcal{G}$ are temporal sources.

5. **Optimal Temporal Spanner**. The property that $\mathcal{G}$ contains a temporal spanner with $2n - 4$ edge labels, where n is the number of vertices in $\mathcal{G}$.

In Section 5, we establish sharp thresholds for the first four properties in RSTGs (see Table 1) and prove an upper bound on the fifth. We conjecture this bound to actually be a sharp threshold as well.

3 Warm-up: 2-hop approach

In this section, we use a simple argument to derive upper bounds on temporal source related properties. The main statement says that for any $p \geq 3\sqrt{\log n/n}$ a fixed vertex v is a temporal source in a random temporal graph from $\mathcal{F}_{n,p}$ a.a.s. The simplicity of this approach comes from the restriction that we consider only temporal paths of length 2. While these bounds are far from optimal, they turn out to be useful for subsequent analyses.

Lemma 3.1. *Let $\alpha \geq 3$ and let $p = \alpha\sqrt{\log n/n}$. Then, for all $n \geq 4$ an arbitrary vertex of $(G,\lambda) \sim \mathcal{F}_{n,p}$ is a temporal source with probability at least $1 - n^{-\alpha^2/4+1}$.*

Proof. Let x be an arbitrary vertex in (G,λ). For two distinct vertices y, z of (G,λ) that are also different from x, we denote by R_z the event that x reaches z, and by S_{yz} we denote the event that x reaches z in exactly two steps via y. Notice that $\mathbb{P}\left[\overline{R_z}\right]$ is the same for every $z \in V(G) \setminus \{x\}$ and we denote this probability by p_1. Similarly, $\mathbb{P}[S_{yz}]$ is the same for all pairs $z, y \in V(G) \setminus \{x\}$, $z \neq y$, and is equal to $p^2/2$. Hence, when denoting by γ the probability that x is a temporal source in (G,λ) and by using the union bound, we have

$$\gamma = 1 - \mathbb{P}\Big(\bigcup_{z \neq x} \overline{R_z}\Big) \geq 1 - \sum_{z \neq x} \mathbb{P}\left[\overline{R_z}\right] = 1 - (n-1)p_1.$$

Furthermore, for every $n \geq 4$ we have

$$\begin{aligned} p_1 = \mathbb{P}\left[\overline{R_z}\right] &\leq \mathbb{P}\Big[\bigcap_{y \neq x,z} \overline{S_{yz}}\Big] = \prod_y \mathbb{P}\left[\overline{S_{yz}}\right] = \mathbb{P}\left[\overline{S_{yz}}\right]^{n-2} \\ &= (1-p^2/2)^{n-2} \leq e^{-\frac{p^2(n-2)}{2}} = \left(\frac{1}{n}\right)^{\alpha^2 \cdot \frac{n-2}{2n}} \leq \left(\frac{1}{n}\right)^{\frac{\alpha^2}{4}}, \end{aligned}$$

where we used the fact that all S_{yz} are independent. Consequently, we derive the desired conclusion

$$\gamma \geq 1 - (n-1)n^{-\alpha^2/4} \geq 1 - n^{-\alpha^2/4+1}.$$

$\square$

If $\alpha = 3$ in the above lemma, we obtain that for $p = 3\sqrt{\log n/n}$ with probability at least $1 - 1/n$ an arbitrary vertex in $(G,\lambda) \sim \mathcal{F}_{n,p}$ is a temporal source. Similarly, if $\alpha = 4$, by applying the union bound, we obtain that for $p = 4\sqrt{\log n/n}$ with probability at least $1 - 1/n^2$ *every* vertex in $(G,\lambda) \sim \mathcal{F}_{n,p}$ is a temporal source, i.e., (G,λ) is temporally connected. However, in the subsequent analysis we will use the following corollary of the lemma obtained by taking $\alpha = \sqrt{\log n}$ and applying the union bound.

Corollary 3.2. *Let $p = \frac{\log n}{\sqrt{n}}$. Then, $(G,\lambda) \sim \mathcal{F}_{n,p}$ is temporally connected with probability at least $1 - n^{-\frac{\log n}{4}+2}$.*

4 Foremost tree evolution

Instead of working with $\mathcal{F}_{n,p}$ directly, it will be often convenient to first draw a temporal graph $\mathcal{G} = (G,\lambda)$ from $\mathcal{F}_{n,1}$ (thus G being a complete graph), and to then consider $\mathcal{G}' = (G',\lambda') = (G,\lambda)_{[0,p]}$ in which all edges with time labels greater than p are deleted. Observe that $G' \sim G_{n,p}$ and that each edge label $\lambda'(e)$ is uniformly distributed on $[0,p]$. In other words, $\mathcal{G}'$ is distributed according to $\mathcal{F}_{n,p}$ up to multiplying all labels by $\frac{1}{p}$.

Consider now the probability that a fixed vertex v reaches another fixed vertex u in $\mathcal{F}_{n,p}$. This is equal to the probability that the temporal subgraph $\mathcal{G}_{[0,p]}$ contains a temporal (v,u)-path P. The latter is again equivalent to the fact that the arrival time of P in $\mathcal{G}$ is at most p. Therefore, the estimation of the parameter p for temporal reachability from v to u can be reduced to the estimation of the arrival time of a *foremost* temporal (v,u)-path in $\mathcal{G}$, i.e., a temporal (v,u)-path that has the smallest arrival time among all temporal (v,u)-paths.

In case of the Temporal Source property, we are interested in the smallest value of p such that a given vertex v is a temporal source in $\mathcal{G}_{[0,p]}$, or equivalently, that any vertex in $\mathcal{G}$ can be reached from v by time p (notice that since $\mathcal{G}$ is a complete temporal graph, every vertex in $\mathcal{G}$ is reachable from v, e.g. by a single-edge path). A minimal temporal subgraph that preserves all foremost paths from v is called *foremost tree for* v, and we will be interested in estimating the smallest p such that $\mathcal{G}$ contains a foremost tree for v with time labels of its edges not exceeding p. We proceed by introducing formally the notion of foremost tree.

Let (T,λ') be a temporal graph, where T is a tree, and let v be a vertex in T. If v is a temporal source (resp. temporal sink) in (T,λ'), then we say that (T,λ') is an *increasing temporal tree* (resp. *decreasing temporal tree*) rooted at v. A temporal subgraph (T,λ') of (G,λ) is a *foremost tree for* v if (a) T is a spanning tree of G; (b) (T,λ') is an increasing temporal tree rooted at v; and (c) for every vertex $u \in V(T)$ the temporal (v,u)-path in (T,λ') is a foremost temporal (v,u)-path in $\mathcal{G}$.

The main purpose of the present section is to analyse temporal and structural properties of a typical foremost tree in $\mathcal{F}_{n,1}$. In Section 4.1 we present an algorithm for constructing a foremost tree for a given temporal source. In Section 4.2 we analyse the execution of the algorithm on a random instance from $\mathcal{F}_{n,1}$ to estimate the speed of growth of the foremost tree. We utilize these results in Section 5 for obtaining sharp thresholds.

4.1 Foremost tree algorithm

We start by presenting an algorithm that given a temporal graph and a source vertex v in the graph constructs a foremost tree for v. The algorithm is similar to Prim's algorithm for a minimum spanning tree in static graphs with the only difference that a next edge to be added to the tree is an edge with the minimum time label *among those that extend the current tree to an increasing temporal tree.*

Algorithm 1 FOREMOST TREE

Input: Simple temporal graph (G,λ); temporal source v in (G,λ).
Output: Foremost tree for v.

$T_0 := (\{v\}, \emptyset)$
for $k := 1$ to $n-1$ **do**
  Let S_k be the set of edges in G with one endpoint in $V(T_{k-1})$ and the other endpoint in $V(G) \setminus V(T_{k-1})$
  $e_k := \arg\min\{\lambda(e) \mid e \in S_k \text{ and } T_{k-1} \cup \{e\} \text{ is an increasing temporal tree rooted at } v\}$
  $T_k := T_{k-1} \cup \{e_k\}$
return (T_{n-1}, λ'), where λ' is the restriction of λ to the edges of T_{n-1}.

In the next lemma we prove the correctness of the algorithm and a simple but crucial property that the time labels of the tree edges monotonically increase as they are added to the tree.

Lemma 4.1. *Let (G,λ) be a simple temporal graph and v be a temporal source in (G,λ). Then* *(i) Algorithm 1 constructs a foremost tree for v in (G, λ);*

(ii) $\lambda(e_1) \leq \lambda(e_2) \leq \ldots \leq \lambda(e_{n-1})$.

Proof. To prove the first claim of the lemma, we will show by induction on the number k of iterations of the **for**-loop that

(a) the tree T_{k-1} can always be extended to an increasing temporal tree, i.e., there always exists an edge e with one endpoint in $V(T_{k-1})$ and the other endpoint in $V(G) \setminus V(T_{k-1})$ such that $T_{k-1} \cup \{e\}$ is an increasing temporal tree;

(b) for every $u \in V(T_k)$ the temporal (v, u)-path in T_k is a foremost (v, u)-path in (G, λ).

The statement is obvious for $k = 1$. Let $1 < k \leq n - 1$ and assume the statement holds for the first $k - 1$ iterations of the **for**-loop. We will show that it also holds for the k-th iteration.

Let u be an arbitrary vertex in $V(G) \setminus V(T_{k-1})$ and let P be a foremost (v, u)-path in (G, λ). Let $e = ab$ be the first edge of P with one endpoint, say a, in T_{k-1} and the other endpoint b not in T_{k-1}. Let also P' be the foremost (v, a)-path in T_{k-1}. Since P' is foremost in (G, λ), the arrival time of P' is not more than the arrival time of the (v, a)-subpath of P. Consequently, the arrival time of P' is not larger than $\lambda(e)$, and therefore $T_{k-1} \cup \{e\}$ is an increasing temporal tree, which proves part (a) of the statement.

Let now $e_k = ab$ be the edge added to T_{k-1} to form T_k, where $a \in V(T_{k-1})$ and $b \notin V(T_{k-1})$. Taking into account the induction hypothesis, to prove part (b) of the statement, we only need to show that the temporal (v, b)-path in T_k is a foremost (v, b)-path in (G, λ). Suppose it is not, and let P be a foremost (v, b)-path in (G, λ). Let e' be the first edge of P with one endpoint in $V(T_{k-1})$ and the other endpoint in $V(T_k)$. By the proof of part (a), we know that $T_{k-1} \cup \{e'\}$ is an increasing temporal tree. Furthermore, since the arrival time of P is less than $\lambda(e_k)$, we have that $\lambda(e') \leq \lambda(e_k)$. But this contradicts the choice of e_k.

To prove the second claim of the lemma, assume that it does not hold and let $k \geq 2$ be the minimum index such that $\lambda(e_k) \leq \lambda(e_{k-1})$. Let $e_k = ab$, where $a \in V(T_{k-1})$ and $b \notin V(T_{k-1})$, and let e_i be the edge of the (v, b)-path in T_k adjacent to e_k. Clearly $\lambda(e_i) \leq \lambda(e_k)$ because T_k is increasing. Hence, there exists j, $i < j \leq k - 1$, such that $\lambda(e_{j-1}) \leq \lambda(e_k) \leq \lambda(e_j)$. Since $i \leq j - 1$, the edge e_i belongs to T_{j-1}, and therefore e_k can extend T_{j-1} to an increasing temporal tree. But this contradicts the choice of e_j at the j-th iteration of the algorithm, as $\lambda(e_k) \leq \lambda(e_j)$. ☐

4.2 Foremost tree growth

The main goal of this section is to estimate time by which a typical foremost tree in $\mathcal{F}_{n,1}$ acquires a given number of vertices. To conduct our analysis, we will consider the execution of Algorithm 1 on a complete random temporal graph $(G, \lambda) \sim \mathcal{F}_{n,1}$ from some fixed vertex v as a random process that reveals the edges of the resulting foremost tree for v one by one in the order they are added to the tree. We define random variables $Y_0^v := 0$ and $Y_k^v := \lambda(e_k^v)$, $k \in [n-1]$, where, following the algorithm, $T_0^v := (\{v\}, \emptyset)$ and for every $k \in [n-1]$

$$
\begin{aligned}
S_k^v &:= V(T_{k-1}^v) \cdot \big(V(G) \setminus V(T_{k-1}^v)\big), \\
e_k^v &:= \arg\min\{\lambda(e) \mid e \in S_k^v \text{ and } T_{k-1}^v \cup \{e\} \text{ is an increasing temporal tree}\}, \\
T_k^v &:= T_{k-1}^v \cup \{e_k^v\}.
\end{aligned}
$$

By definition, Y_k^v is the earliest time when the foremost tree for v contains exactly k edges, or equivalently the earliest time by which v can reach $k + 1$ vertices. For $k \in [n-1]$, let X_k^v be a random variable equal to $Y_k^v - Y_{k-1}^v$, i.e., to the *waiting time* between the edges e_{k-1}^v and e_k^v. Clearly, we have

$$Y_k^v = \sum_{i=1}^{k} X_i^v$$

for every $k \in [n-1]$.

The main objects of our analysis are random variables $X_1^v, X_2^v, \dots, X_{n-1}^v$ and $Y_1^v, Y_2^v, \dots, Y_{n-1}^v$. We will also study the behaviour of their *truncated* versions, which are convenient in the applications: For $k \in [n-1]$, let $\widehat{X}_k^v := \min\{X_k^v, c_k\}$ and $\widehat{Y}_k^v := \sum_{i=1}^k \widehat{X}_i^v$, where

$$c_k := \frac{2\log(\min\{k, n-k\}) + \log\log n}{k(n-k)}.$$

The values of c_k are chosen in such a way that on the one hand they are sufficiently small, and on the other hand they are large enough to guarantee that the truncated variables coincide with their original versions with high probability. This is formalized in the following lemma.

Lemma 4.2 (Properties of c_k)**.** *We have*

(i) $\sum_{i=1}^{n-1} c_i^2 \leq \frac{64(\log\log n)^2}{n^2}$*;*

(ii) for a fixed vertex v, with probability at least $1 - 4/\log n$ the equality $\widehat{X}_k^v = X_k^v$ holds for every $k \in [n-1]$.

Proof. We start by proving the first part of the lemma.

$$\begin{aligned}
\sum_{i=1}^{n-1} c_i^2 &= \sum_{i=1}^{n-1} \frac{\big(2\log(\min\{i,n-i\}) + \log\log n\big)^2}{(i(n-i))^2} \leq \sum_{i=1}^{n-1} \frac{\big(2\log(\min\{i,n-i\}) \cdot \log\log n\big)^2}{(i(n-i))^2} \\
&\leq 4(\log\log n)^2 \left(\sum_{i=1}^{\lfloor n/2 \rfloor} \frac{(\log i)^2}{i^2 (n/2)^2} + \sum_{i=\lfloor n/2 \rfloor + 1}^{n-1} \frac{(\log(n-i))^2}{(n/2)^2 (n-i)^2} \right) \\
&\leq \frac{32(\log\log n)^2}{n^2} \sum_{i=1}^{\infty} \frac{(\log i)^2}{i^2} \leq \frac{64(\log\log n)^2}{n^2}.
\end{aligned}$$

To prove the second part, we observe that

$$\begin{aligned}
\mathbb{P}\big[\widehat{X}_k^v \neq X_k^v\big] = \mathbb{P}\big[X_k^v > c_k\big] &\leq (1-c_k)^{|S_k^v|} = (1-c_k)^{k(n-k)} \\
&\leq e^{-k(n-k)c_k} = \frac{1}{(\min\{k,n-k\})^2 \log n},
\end{aligned} \tag{1}$$

and therefore

$$\begin{aligned}
\sum_{k=1}^{n-1} \mathbb{P}\big[\widehat{X}_k^v \neq X_k^v\big] &\leq \sum_{k=1}^{n-1} \frac{1}{(\min\{k,n-k\})^2 \log n} \\
&= \frac{1}{\log n} \left(\sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k^2} + \sum_{k=\lfloor n/2 \rfloor + 1}^{n-1} \frac{1}{(n-k)^2} \right) \\
&\leq \frac{2}{\log n} \sum_{k=1}^{\infty} \frac{1}{k^2} \leq \frac{4}{\log n},
\end{aligned}$$

which implies the result. $\square$

We will now estimate the expected time between the moments of exposing two consecutive edges of the foremost tree. More specifically, we will bound the expected values of X_k^v and $\widehat{X}_k^v$ given the information revealed by the process in the first $k-1$ steps. For every $k \in [n-1]$, let $\mathcal{A}_k^v$ be the σ-algebra generated by the information revealed at the first k iteration of the algorithm starting at v, i.e., by the knowledge of the first k edges $e_1^v, e_2^v, \ldots, e_k^v$ and their time labels $Y_1^v, Y_2^v, \ldots, Y_k^v$. Let also $\mathcal{A}_0^v$ be the trivial σ-algebra. Then, we have the following

Lemma 4.3. *For a vertex v and every $k \in [n-1]$ we have*

(i)
$$\frac{1 - Y_{k-1}^v}{k(n-k)+1} \leq \mathbb{E}[X_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)+1};$$

(ii)
$$(1 - 1/\log n) \cdot \frac{1 - Y_{k-1}^v}{k(n-k)+1} \leq \mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)+1}.$$

Proof. For every $k \in [n-1]$ we define the function $w_k^v : S_k^v \to [0,1]$ as follows

$$w_k^v(e) = \begin{cases} \lambda(e) - Y_{k-1}^v, & \lambda(e) \geq Y_{k-1}^v \\ \lambda(e) - Y_{k-1}^v + 1, & \lambda(e) < Y_{k-1}^v. \end{cases}$$

Notice that for any two edges $e, f \in S_k^v$ such that $\lambda(f) < Y_{k-1}^v \leq \lambda(e)$ we have $w_k^v(e) < w_k^v(f)$. This together with Lemma 4.1 (ii) implies that e_k^v is exactly the edge on which the minimum of w_k^v is attained, that is,

$$e_k^v = \arg\min\{w_k^v(e) \mid e \in S_k^v\}, \tag{2}$$

and therefore, for every $k \in [n-1]$,

$$X_k^v = \min\{w_k^v(e) \mid e \in S_k^v\}. \tag{3}$$

Observe that upon exposure of edge e_k^v at step k of the algorithm, we reveal some information about the time labels of the other edges in S_k^v. More precisely, we learn that these time labels are not contained in the interval $[Y_{k-1}^v, Y_k^v]$. Thus, if we inductively define the admissible range of $\lambda(e)$, $e \in S_k^v$, as

$$I_k^v(e) := \begin{cases} I_{k-1}^v(e) \setminus [Y_{k-2}^v, Y_{k-1}^v], & e \in S_k^v \cap S_{k-1}^v \\ [0,1], & e \in S_k^v \setminus S_{k-1}^v \end{cases}$$

then $\lambda(e)$ conditioned on $\mathcal{A}_{k-1}^v$ is uniformly distributed on $I_k^v(e)$. Let $\ell(e)$ be the unique index with $e \in S_\ell^v \setminus S_{\ell-1}^v$, then we have $I_k^v(e) = [0,1] \setminus [Y_{\ell-1}^v, Y_{k-1}^v]$. It follows that $w_k^v(e)$ is uniformly distributed on its admissible range

$$J_k^v(e) := [0, Y_{\ell-1}^v - Y_{k-1}^v + 1]$$

and clearly

$$[0, 1 - Y_{k-1}^v] \subseteq J_k^v(e) \subseteq [0,1]\,. \tag{4}$$

Note that X_k^v is a minimum of $k(n-k)$ independent random variables $w_k^v(e), e \in S_k^v$, where for every edge $e \in S_k^v$ the value $w_k^v(e)$ is distributed uniformly on its own admissible range $J_k^v(e)$. Let X_k' be the minimum of $k(n-k)$ independent random variables distributed uniformly in $\left[0, 1 - Y_{k-1}^v\right]$, and X_k'' be the minimum of $k(n-k)$ independent random variables distributed uniformly in $[0,1]$. Then, the inclusion (4) implies

$$\mathbb{P}[X_k' \geq t] \leq \mathbb{P}[X_k^v \geq t \mid A_{k-1}^v] \leq \mathbb{P}[X_k'' \geq t],$$

where A_{k-1}^v is the event that specific edges $e_1^v, e_2^v, \ldots, e_{k-1}^v$ with time labels $Y_1^v, Y_2^v, \ldots, Y_{k-1}^v$ are revealed in the first $k-1$ steps of the process. Therefore, by integrating and noting that the expected value of the minimum of m independent variables distributed uniformly on $[0, a]$ is equal to $\frac{a}{m+1}$, we obtain

$$\frac{1 - Y_{k-1}^v}{k(n-k)+1} = \mathbb{E}[X'_k] \leq \mathbb{E}[X_k^v \mid \mathcal{A}_{k-1}] \leq \mathbb{E}[X''_k] = \frac{1}{k(n-k)+1}. \tag{5}$$

To prove the second part of the lemma we first note that by definition $\widehat{X}_k^v \leq X_k^v$, and hence

$$\mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v] \leq \mathbb{E}[X_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)+1}.$$

Therefore, it remains to show the lower bound on $\mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v]$. For convenience, let us denote $M_k := k(n-k)$. Then, we have

$$\begin{aligned}
\mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v] &= \int_0^\infty \mathbb{P}[\widehat{X}_k^v \geq t \mid A_{k-1}^v]\,\mathrm{d}t = \int_0^{c_k} \mathbb{P}[\widehat{X}_k^v \geq t \mid A_{k-1}^v]\,\mathrm{d}t \\
&= \int_0^{c_k} \mathbb{P}[X_k^v \geq t \mid A_{k-1}^v]\,\mathrm{d}t \geq \int_0^{c_k} \mathbb{P}[X'_k \geq t]\,\mathrm{d}t \\
&= \int_0^{c_k} \left(1 - \frac{t}{1 - Y_{k-1}^v}\right)^{M_k} \mathrm{d}t \\
&= \frac{(1 - Y_{k-1}^v) + (c_k - (1 - Y_{k-1}^v))\left(1 - \frac{c_k}{1 - Y_{k-1}^v}\right)^{M_k}}{M_k + 1} \\
&\geq \frac{(1 - Y_{k-1}^v) - (1 - Y_{k-1}^v)(1 - c_k)^{M_k}}{M_k + 1} \\
&= \frac{(1 - Y_{k-1}^v)\left(1 - (1 - c_k)^{M_k}\right)}{M_k + 1},
\end{aligned}$$

and the desired bound follows from the fact that $(1 - c_k)^{k(n-k)} \leq \frac{1}{\log n}$ (see Eq. (1)). $\square$

Recall that, by Corollary 3.2, $\mathcal{F}_{n,\,(\log n)/\sqrt{n}}$ is temporally connected with probability at least $1 - n^{-\frac{\log n}{4}+2}$, in which case the bound $Y_k^v \leq Y_{n-1}^v \leq \frac{\log n}{\sqrt{n}}$ holds for every v and $k \in [n-1]$. Thus, we derive the following

Corollary 4.4. *With probability at least $1 - n^{-\frac{\log n}{4}+2}$, for every vertex v and every $k \in [n-1]$ we have*

(i) $$\left(1 - \frac{\log n}{\sqrt{n}}\right) \cdot \frac{1}{k(n-k)+1} \leq \mathbb{E}[X_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)+1};$$

(ii) $$\left(1 - \frac{2}{\log n}\right) \cdot \frac{1}{k(n-k)+1} \leq \mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)+1}.$$

Next, we will bound the deviation of the truncated time of the moment when the foremost tree acquires k edges from the expected value of accumulated truncated waiting times between the consecutive edges in the sequence of the first k edges of the tree. For this we require the following standard inequality by Azuma.

Theorem 4.5 (Azuma's inequality [4])**.** *Let $Z_0, Z_1, \ldots, Z_n$ be a martingale with respect to a filtration $\{\emptyset, \Omega\} = \mathcal{A}_0 \subset \mathcal{A}_1 \subset \ldots \subset \mathcal{A}_n$. Let also $c_1, c_2, \ldots, c_n$ be non-negative numbers such that $\sum_{i=1}^{n} \mathbb{P}\left[|Z_i - Z_{i-1}| \geq c_i\right] = 0$. Then*

$$\mathbb{P}\big[|Z_n - Z_0| \geq \mu\big] \leq 2\exp\left(\frac{-\mu^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

Lemma 4.6. *For a fixed vertex v, with probability at least $1 - n^{-\sqrt{\log n}-1}$ the inequality*

$$\left|\widehat{Y}_k^v - \sum_{i=1}^{k} \mathbb{E}\left[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v\right]\right| \leq \frac{(\log n)^{0.8}}{n}$$

holds for all $k \in [n-1]$.

Proof. Let us fix $k \in [n-1]$ and define a martingale $Z_0^v, Z_1^v, \ldots, Z_k^v$ with $Z_0^v := 0$ and

$$Z_s^v := \widehat{Y}_s^v - \sum_{i=1}^{s} \mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v] = \sum_{i=1}^{s} \widehat{X}_i^v - \sum_{i=1}^{s} \mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v],$$

for $s \in [k-1]$.

Since $0 \leq \widehat{X}_i^v \leq c_i$, we have $0 \leq \mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v] \leq c_i$, and therefore

$$\mathbb{P}\left[\left|Z_i^v - Z_{i-1}^v\right| > c_i\right] = \mathbb{P}\left[\left|\widehat{X}_i^v - E[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v]\right| > c_i\right] = 0$$

holds for every $i \in [k]$. Furthermore, by Lemma 4.2 (i)

$$\sum_{i=1}^{k} c_i^2 \leq \sum_{i=1}^{n-1} c_i^2 \leq \frac{64(\log\log n)^2}{n^2},$$

and hence applying Azuma's inequality (Theorem 4.5), we obtain that for sufficiently large n

$$\begin{aligned}
\mathbb{P}\left[|Z_k^v| \geq \frac{(\log n)^{0.8}}{n}\right] &\leq 2\exp\left(\frac{-(\log n)^{1.6}}{n^2} \cdot \frac{1}{2\sum_{i=1}^{k} c_k^2}\right) \\
&\leq 2\exp\left(\frac{-(\log n)^{1.6}}{n^2} \cdot \frac{n^2}{128(\log\log n)^2}\right) \\
&= 2\exp\left(\frac{-(\log n)^{1.6}}{128(\log\log n)^2}\right) \\
&\leq \exp\left(-(\log n)^{1.55}\right) \\
&\leq n^{-\sqrt{\log n}-2}.
\end{aligned}$$

The latter inequality together with the union bound over all $k \in [n-1]$ imply the desired result. ☐

Finally, having Corollary 4.4 and Lemma 4.6 at hand we are ready to prove the main result of this section.

Theorem 4.7. *With probability at least $1 - 2n^{-\sqrt{\log n}}$, for every vertex v and $k \in [n-1]$ we have*

$$\left|\widehat{Y}_k^v - \sum_{i=1}^{k} \frac{1}{i(n-i)+1}\right| \leq \frac{2(\log n)^{0.8}}{n}.$$

Proof. By Corollary 4.4 (ii), with probability at least $1-n^{-\frac{\log n}{4}+2}$, for every vertex v and every $k \in [n-1]$ we have

$$-\sum_{i=1}^{k}\mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v] - \frac{2}{\log n}\cdot\sum_{i=1}^{k}\frac{1}{i(n-i)+1} \leq -\sum_{i=1}^{k}\frac{1}{i(n-i)+1} \leq -\sum_{i=1}^{k}\mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v]. \quad (6)$$

Similarly, by the union bound and Lemma 4.6, with probability at least $1-n^{-\sqrt{\log n}}$ for every vertex v and every $k \in [n-1]$ we have

$$\sum_{i=1}^{k}\mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v] - \frac{(\log n)^{0.8}}{n} \leq \widehat{Y}_k^v \leq \sum_{i=1}^{k}\mathbb{E}[\widehat{X}_i^v \mid \mathcal{A}_{i-1}^v] + \frac{(\log n)^{0.8}}{n}. \quad (7)$$

Hence, summing up (6) and (7), we conclude that with probability at least $1-2n^{-\sqrt{\log n}}$ for every vertex v and every $k \in [n-1]$

$$-\frac{(\log n)^{0.8}}{n} - \frac{2}{\log n}\cdot\sum_{i=1}^{k}\frac{1}{i(n-i)+1} \leq \widehat{Y}_k^v - \sum_{i=1}^{k}\frac{1}{i(n-i)+1} \leq \frac{(\log n)^{0.8}}{n},$$

which implies the result after noticing that

$$\begin{aligned}\frac{2}{\log n}\cdot\sum_{i=1}^{k}\frac{1}{i(n-i)+1} &\leq \frac{2}{\log n}\cdot\sum_{i=1}^{n-1}\frac{1}{i(n-i)} = \frac{2}{n\cdot\log n}\cdot\sum_{i=1}^{n-1}\left(\frac{1}{i}+\frac{1}{n-i}\right)\\ &= \frac{4}{n\cdot\log n}\cdot\sum_{i=1}^{n-1}\frac{1}{i} \leq \frac{6}{n} \in o\left(\frac{(\log n)^{0.8}}{n}\right).\end{aligned}$$

□

Corollary 4.8. *With probability at least $1-2n^{-\sqrt{\log n}}$, for every vertex v we have*

$$\left|\widehat{Y}_{n-1}^v - \frac{2\log n}{n}\right| \leq \frac{3(\log n)^{0.8}}{n}.$$

Proof. Using Theorem 4.7 and the equality

$$\sum_{i=1}^{n-1}\frac{1}{i(n-i)} = \frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{1}{i}+\frac{1}{n-i}\right) = \frac{2}{n}\sum_{i=1}^{n-1}\frac{1}{i} = \frac{2\log n + O(1)}{n},$$

we conclude that with probability at least $1-2n^{-\sqrt{\log n}}$ the inequality

$$\begin{aligned}\left|\widehat{Y}_{n-1}^v - \frac{2\log n}{n}\right| &\leq \left|\widehat{Y}_{n-1}^v - \sum_{i=1}^{n-1}\frac{1}{i(n-i)}\right| + \left|\sum_{i=1}^{n-1}\frac{1}{i(n-i)} - \frac{2\log n}{n}\right|\\ &\leq \frac{2(\log n)^{0.8}}{n} + \frac{O(1)}{n} \leq \frac{3(\log n)^{0.8}}{n}\end{aligned}$$

holds for every vertex v. □

5 Sharp thresholds for temporal graph properties

In this section, we apply the results obtained in Section 4.2 to establish sharp thresholds for Point-to-point Reachability, First Temporal Source, Temporal Source, and Temporal Connectivity properties. We recall that our general strategy for obtaining a sharp threshold p_0 for a certain property in the model $\mathcal{F}_{n,p}$ is to show that the property *does not* hold in a random temporal complete graph $\mathcal{F}_{n,1}$ before time p_0, and holds after time p_0 a.a.s. Therefore, in the proofs, if we do not specify explicitly the model from which a graph under consideration comes from, we assume that it comes from $\mathcal{F}_{n,1}$.

5.1 Point-to-point Reachability

Recall that Point-to-point Reachability is the property that for a fixed pair of vertices there exists a temporal path from the first vertex to the second one. In this section we establish a sharp threshold for this property.

Theorem 5.1. *The function $\frac{\log n}{n}$ is a sharp threshold for Point-to-point Reachability. More specifically, for any sufficiently large n, for two fixed distinct vertices v and u in $\mathcal{G} \sim \mathcal{F}_{n,p}$*

(i) there is no *temporal (v,u)-path in $\mathcal{G}$ with probability at least $1 - \frac{6}{\log n}$, if $p < \frac{\log n}{n} - \varepsilon(n)$; and*

(ii) there is a temporal (v,u)-path in $\mathcal{G}$ with probability at least $1 - \frac{6}{\log n}$, if $p > \frac{\log n}{n} + \varepsilon(n)$,

where $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.

Proof. We set $k := \frac{n-1}{2\log n}$ and $k' := n-1-k$ and apply Theorem 4.7 to estimate $\widehat{Y}^v_k$ and $\widehat{Y}^v_{k'}$. We obtain that with probability at least $1 - 2n^{-\sqrt{\log n}} > 1 - \frac{1}{\log n}$ the following inequalities hold:

$$\widehat{Y}^v_k \geq -\frac{2(\log n)^{0.8}}{n} + \sum_{i=1}^{k} \frac{1}{i(n-i)}; \qquad \widehat{Y}^v_{k'} \leq \frac{2(\log n)^{0.8}}{n} + \sum_{i=1}^{k'} \frac{1}{i(n-i)}.$$

For all large enough n, the first inequality implies

$$\begin{aligned}
\widehat{Y}^v_k &\geq -\frac{2(\log n)^{0.8}}{n} + \sum_{i=1}^{k} \frac{1}{i(n-i)} \\
&= -\frac{2(\log n)^{0.8}}{n} + \frac{1}{n}\sum_{i=1}^{k} \left(\frac{1}{i} + \frac{1}{n-i}\right) \\
&= -\frac{2(\log n)^{0.8}}{n} + \frac{\log k + \log n - \log(n-k) + O(1)}{n} \\
&= -\frac{2(\log n)^{0.8}}{n} + \frac{\log n - O(\log\log n)}{n} \\
&\geq \frac{\log n - 3(\log n)^{0.8}}{n},
\end{aligned}$$

where we used the approximation for the harmonic series. For the same reasons, the second inequality implies

$$\widehat{Y}^v_{k'} \leq \frac{\log n + 3(\log n)^{0.8}}{n}.$$

Furthermore, by Lemma 4.2 (ii) with probability at least $1 - \frac{4}{\log n}$ we have $\widehat{X}_j^v = X_j^v$ for all $j \in [n-1]$, thus in particular $\widehat{Y}_k^v = Y_k^v$ and $\widehat{Y}_{k'}^v = Y_{k'}^v$. By symmetry, the probability of u to be among the first k or the last k vertices to become reachable from v is equal to $\frac{2k}{n-1} = \frac{1}{\log n}$. Hence, since the first k vertices become reachable at the time Y_k^v, and all but the last k vertices at the time $Y_{k'}^v$, by the union bound, with probability at least $1 - \frac{6}{\log n}$, vertex u is not reachable from v by the time $Y_k^v = \widehat{Y}_k^v \geq \frac{\log n}{n} - \frac{3(\log n)^{0.8}}{n}$, and u is reachable from v by the time $Y_{k'}^v = \widehat{Y}_{k'}^v \leq \frac{\log n}{n} + \frac{3(\log n)^{0.8}}{n}$, which implies the result. $\square$

5.2 Temporal Source and First Temporal Source

In this section, we first establish a sharp threshold for Temporal Source, i.e., for the property that a fixed vertex is a temporal source. Then, we show that, quite surprisingly, the same function happens to be a sharp threshold for the property of having at least one temporal source in the graphs, i.e., for First Temporal Source.

Before proceeding we make the simple but subsequently useful observation that the probability of a vertex v of $\mathcal{F}_{n,p}$ being a temporal source is equal to the probability of v being a temporal sink. This follows from the fact that v is a temporal source in (G, λ) if and only if it is a temporal sink in (G, λ'), where $\lambda'(e) := 1 - \lambda(e)$ for every edge of G. Therefore, the next two theorems remain valid if every occurrence of the word "source" is replaced by "sink". With a slight abuse of formalities we will refer to Theorem 5.2 and Theorem 5.3 in both "source" and "sink" scenarios.

Theorem 5.2. *The function $\frac{2\log n}{n}$ is a sharp threshold for Temporal Source. More specifically, for any sufficiently large n, a fixed vertex v in $\mathcal{G} \sim \mathcal{F}_{n,p}$*

(i) is not *a temporal source with probability at least $1 - \frac{5}{\log n}$, if $p < \frac{2\log n}{n} - \varepsilon(n)$;*

(ii) is a temporal source with probability at least $1 - \frac{5}{\log n}$, if $p > \frac{2\log n}{n} + \varepsilon(n)$,

where $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.

Proof. By Corollary 4.8, the inequality $|\widehat{Y}_{n-1}^v - \frac{2\log n}{n}| \leq \frac{3(\log n)^{0.8}}{n}$ holds with probability at least $1 - 2n^{-\sqrt{\log n}}$. Furthermore, by Lemma 4.2 (ii), the equality $\widehat{Y}_{n-1}^v = Y_{n-1}^v$ holds with probability at least $1 - \frac{4}{\log n}$. Hence, by the union bound, the inequality $|Y_{n-1}^v - \frac{2\log n}{n}| \leq \frac{3(\log n)^{0.8}}{n}$ holds with probability at least $1 - \frac{5}{\log n}$. This implies the result, since Y_{n-1}^v is exactly the time when v becomes temporal source. $\square$

Theorem 5.3. *The function $\frac{2\log n}{n}$ is a sharp threshold for First Temporal Source. More specifically, for any sufficiently large n, a random temporal graph $\mathcal{G} \sim \mathcal{F}_{n,p}$*

(i) does not *contain a temporal source with probability at least $1 - 2n^{-\sqrt{\log n}}$, if $p < \frac{2\log n}{n} - \varepsilon(n)$;*

(ii) contains a temporal source with probability at least $1 - \frac{5}{\log n}$, if $p > \frac{2\log n}{n} + \varepsilon(n)$,

where $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.

Proof. We observe that the upper bound of $\frac{2\log n}{n} + \varepsilon(n)$ on the threshold follows from Theorem 5.2, and we only need to show the lower bound. Recall that by definition we always have $\widehat{Y}_{n-1}^v \leq Y_{n-1}^v$. Therefore, by Corollary 4.8, with probability at least $1 - 2n^{-\sqrt{\log n}}$ the inequality $Y_{n-1}^v \geq \widehat{Y}_{n-1}^v \geq \frac{2\log n}{n} - \frac{3(\log n)^{0.8}}{n}$ holds for all vertices v. In other words, with probability at least $1 - 2n^{-\sqrt{\log n}}$ no vertex is a temporal source before the time $\frac{2\log n}{n} - \frac{3(\log n)^{0.8}}{n}$. $\square$

5.3 Temporal Connectivity

In this section, we establish a sharp threshold for Temporal Connectivity.

Theorem 5.4. *The function $\frac{3\log n}{n}$ is a sharp threshold for Temporal Connectivity. More specifically, for any sufficiently large n, a random temporal graph in $\mathcal{F}_{n,p}$*

(i) is not *temporally connected a.a.s., if $p < \frac{3\log n}{n} - \varepsilon(n)$, where $\varepsilon(n) = \frac{6(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$;*

(ii) is temporally connected a.a.s., if $p > \frac{3\log n}{n} + \varepsilon(n)$, where $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.

We split the proof of Theorem 5.4 in two parts. Note that a temporal graph is temporally connected if and only if each of its vertices is a temporal sink. Our strategy it to show that on the one hand before the time $\frac{3\log n}{n}$ at least one vertex in a random temporal complete graph $\mathcal{G} \sim \mathcal{F}_{n,1}$ is not a temporal sink a.a.s. (Lemma 5.5), and on the other hand after the time $\frac{3\log n}{n}$ all vertices in $\mathcal{G}$ are temporal sinks a.a.s. (Lemma 5.6).

Lemma 5.5. *Let $p < \frac{3\log n}{n} - \varepsilon(n)$, where $\varepsilon(n) = \frac{6(\log n)^{0.8}}{n}$, and let $\mathcal{G} \sim \mathcal{F}_{n,1}$. Then, a.a.s. the temporal graph $\mathcal{G}_{[0,p]}$ contains at least one vertex which is not a temporal sink.*

Proof. Let q be the time at which the first vertex in $\mathcal{G}$ becomes a temporal sink, i.e.,

$$q := \min\{q' \mid \mathcal{G}_{[0,q']} \text{ has a temporal sink}\}.$$

By Theorem 5.3, $q \geq \frac{2\log n}{n} - \frac{3(\log n)^{0.8}}{n} = \frac{2\log n}{n} - \frac{\varepsilon}{2}$ a.a.s. Note that $\mathcal{G}_{[0,q]}$ has at most two temporal sinks, since the addition of a single edge can not turn more than two vertices into temporal sinks. Furthermore, in order for a vertex which is not a temporal sink at time q to become a temporal sink it has to a acquire at least one edge incident to it after time q. Thus, to prove the lemma, we will show that the underlying graph H of $\mathcal{G}_{[q,p]}$ has at least 3 isolated vertices, and hence at least one of these vertices is not a temporal sink in $\mathcal{G}_{[0,p]}$.

For the purpose of this, note that any pair of vertices forms an edge in H with probability $u := \frac{p-q}{1-q}$, unless it is an edge in $\mathcal{G}_{[0,q]}$. Thus, graph H is distributed the same as taking an element of $G_{n,u}$ and then deleting all edges contained in $\mathcal{G}_{[0,q]}$.

As $u = \frac{p-q}{1-q}$ is maximized when q is minimal, we have for sufficiently large n that

$$\begin{aligned} u = \frac{p-q}{1-q} &\leq \frac{3\frac{\log n}{n} - \varepsilon - 2\frac{\log n}{n} + \varepsilon/2}{1 - 2\frac{\log n}{n} + \varepsilon/2} < \frac{\frac{\log n}{n} - \frac{\varepsilon}{2}}{1 - 2\frac{\log n}{n}} \\ &= \frac{1}{n} \cdot \frac{\log n - 3(\log n)^{0.8}}{1 - 2\frac{\log n}{n}} = \frac{1}{n}\left(\log n - \frac{3(\log n)^{0.8} - 2\frac{(\log n)^2}{n}}{1 - 2\frac{\log n}{n}}\right) \\ &< \frac{1}{n}\left(\log n - \frac{2(\log n)^{0.8}}{1 - 2\frac{\log n}{n}}\right) < \frac{\log n - (\log n)^{0.8}}{n}. \end{aligned}$$

It is known that $G_{n,u}$ contains a.a.s. more than two isolated vertices if $\lim_{n\to\infty} n(1-u)^{n-1} = \infty$ [6, Theorem 3.1(ii)]. In order to show the latter, we first evaluate

$$\tau := \log n + n \cdot \log\left(1 - \frac{\log n - (\log n)^{0.8}}{n}\right) = (\log n)^{0.8} - \mathcal{O}\left(\frac{(\log n)^2}{n}\right) \xrightarrow{n\to\infty} \infty$$

by means of Maclaurin expansion. Now, using this, we derive

$$\begin{aligned} n(1-u)^{n-1} &\geq n\left(1-\frac{\log n-(\log n)^{0.8}}{n}\right)^{n-1} \\ &= n\cdot\exp\left((n-1)\log\left(1-\frac{\log n-(\log n)^{0.8}}{n}\right)\right) \\ &\geq e^{\tau} \xrightarrow{n\to\infty} \infty\,. \end{aligned}$$

Thus, $G_{n,u}$, and therefore H, contains at least three isolated vertices a.a.s., as required. $\square$

Lemma 5.6. *Let* $p > \frac{3\log n}{n} + \varepsilon(n)$*, where* $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n}$*, and let* $\mathcal{G} \sim \mathcal{F}_{n,1}$*. Then, a.a.s. every vertex in* $\mathcal{G}_{[0,p]}$ *is a temporal sink.*

Proof. Let $q := \frac{2\log n}{n} + \varepsilon(n)$ and $r := \frac{n}{\log n}\log\log\log n$. It follows from Theorem 5.2 that the probability for an arbitrary vertex *not* to be temporal sink in $\mathcal{G}_{[0,q]}$ is at most $\frac{5}{\log n}$. Hence, the expected number of vertices that are not temporal sinks in $\mathcal{G}_{[0,q]}$ is at most $\frac{5n}{\log n}$, and therefore, by Markov's inequality, the probability of $\mathcal{G}_{[0,q]}$ having more than r vertices that are not temporal sinks is at most $\frac{5n}{\log n}\cdot\frac{\log n}{n\cdot\log\log\log n} = \frac{5}{\log\log\log n} \in o(1)$. We will show that no more than $\frac{\log n}{n}$ extra time is required for these at most r remaining vertices to become temporal sinks.

More formally, let $S \subseteq V$ be the set of vertices that are temporal sinks in $\mathcal{G}_{[0,q]}$. Observe that if at some time $t > q$ a vertex $w \in V \setminus S$ acquires an edge that connects w with a temporal sink in S, then w becomes a temporal sink no later than time t (notice, that w can become a temporal sink before time t). Therefore, to prove the lemma, we will show that a.a.s. by time $q + \frac{\log n}{n}$ every vertex in $V \setminus S$ acquires such an edge.

First, note that the underlying graph $G_{[0,q]}$ of $\mathcal{G}_{[0,q]}$ is distributed according to $G_{n,q}$. Hence, by applying [6, Corollary 3.13] (with the function $\omega(n) = \log(n)/2$), we obtain that the maximum degree $\Delta(G_{[0,q]})$ of $G_{[0,q]}$ a.a.s. satisfies

$$\begin{aligned} \Delta(G_{[0,q]}) &\leq qn + \sqrt{2q(1-q)n\cdot\log n} + \frac{\log(n)}{2}\sqrt{\frac{q(1-q)n}{\log n}} \\ &< 3\log(n) + \sqrt{6}\cdot\log(n) + \frac{\sqrt{3}}{2}\cdot\log(n) < 7\log(n) \end{aligned}$$

where we use that $\varepsilon(n) < \frac{\log(n)}{n}$ for sufficiently large n. For every vertex $w \in V \setminus S$ define

$$C_w := \{vw \mid v \in S \text{ and } \lambda(vw) > q\}$$

as the set of edges connecting w to S after time q. By the above, a.a.s. for every w we have that $|C_w| \geq n-1-r-7\log(n) \geq n-2r =: d$. Let T_w be the *waiting time* for the first of these edges to appear, i.e., $T_w := \min\{\lambda(vw) - q \mid vw \in C_w\}$. Note that the time labels of the edges in C_w are independently and uniformly distributed on the interval $I = (q,1]$. Thus, we have $\mathbb{P}[T_w > x] \leq (1-x)^d$ for any $0 \leq x \leq 1-q$.

Now let T be the waiting time of the last vertex in $V \setminus S$ to become adjacent to at least one of the temporal sinks in S, i.e., $T = \max\{T_w \mid w \in V \setminus S\}$. Because the waiting times $T_w, w \in V \setminus S$ are all independent, we have

$$\mathbb{P}[T > x] = 1 - \prod_w \mathbb{P}[T_w < x] \leq 1 - (1-(1-x)^d)^r.$$

In the rest of the proof we show that $\lim_{n\to\infty} \mathbb{P}[T > \frac{\log n}{n}] = 0$. For this purpose, consider

$$\begin{aligned}
\tau := \mathbb{P}\left[T_w > \frac{\log n}{n}\right] &\le \left(1 - \frac{\log n}{n}\right)^d \\
&= \left(1 - \frac{\log n}{n}\right)^{\frac{n}{\log n}(\log n - 2\log\log\log n)} \\
&\le e^{-(\log n - 2\log\log\log n)} \\
&= \frac{(\log\log n)^2}{n} < 1\,.
\end{aligned}$$

Using this, we derive

$$\begin{aligned}
0 \ge r \cdot \log(1-\tau) \ge r \cdot \log\left(1 - \frac{(\log\log n)^2}{n}\right) \\
&\ge -2r\frac{(\log\log n)^2}{n} \quad [\text{since } (\log\log n)^2/n < 0.7, \text{ assuming } n \ge 2] \\
&= -2\frac{(\log\log n)^2 \cdot \log\log\log n}{\log n} \xrightarrow{n\to\infty} 0\,.
\end{aligned}$$

Thus, we have $\lim_{n\to\infty} r \cdot \log(1-\tau) = 0$ and therefore

$$\mathbb{P}\left[T > \frac{\log n}{n}\right] \le 1 - (1-\tau)^r = 1 - \exp\left(r \cdot \log(1-\tau)\right) \xrightarrow{n\to\infty} 1 - e^0 = 0\,.$$

□

5.4 Existence of optimal spanners

In this section, we discuss the existence of *optimal spanners*, i.e., spanners consisting of exactly $2n-4$ edge appearances. It is known that no spanner can have fewer than $2n-4$ edge appearances even for graphs with multiple appearances per edge [9]. However, a temporal graph can be temporally connected without admitting an optimal spanner; for instance, no temporal graph whose underlying graph does not contain a 4-cycle contains an optimal spanner [9].

Since existence of a spanner assumes temporal connectivity, Lemma 5.5 implies that $\mathcal{G} \sim \mathcal{F}_{n,p}$ has no optimal spanner a.a.s. whenever $p < (3-o(1))\frac{\log n}{n}$. The main goal of the present section is to show that $\mathcal{G} \sim \mathcal{F}_{n,p}$ contains an optimal spanner a.a.s. whenever $p > (4+o(1))\frac{\log n}{n}$.

The proof of this result relies upon a construction of a *pair of temporal trees*. The basic idea is to pick two trees rooted at the same vertex (referred to as the *pivot*) such that all graph vertices can reach the pivot before some time t using only the edges of one of the trees, and the pivot can reach all graph vertices after time t using only the edges of the other tree. Clearly, the union of these two trees is a spanner with $2n-2$ edges. This construction can be improved to $2n-4$ by replacing the pivot vertex by a pivot 4-cycle.

In the proof we use the notion of "reverse foremost tree" to w, by which we mean a decreasing temporal tree rooted at w that one obtains by running Algorithm 1 from vertex w backwards in time.

Theorem 5.7. *A random temporal graph in $\mathcal{F}_{n,p}$ a.a.s. has an optimal spanner if $p \ge 4\frac{\log n}{n} + \varepsilon(n)$ where $\varepsilon(n) := \frac{16(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.*

Proof. Assume $n \geq 8$ and set $\varepsilon_0(n) := \frac{\varepsilon(n)}{4} = \frac{4(\log n)^{0.8}}{n} > \frac{3(\log(n-3))^{0.8}}{n-3} > \frac{\log\log n}{n}$. Let also $\mathcal{G} \sim \mathcal{F}_{n,1}$. Our goal is to show that $\mathcal{G}_{[0,p]}$ has an optimal spanner a.a.s. For this purpose, set

$$p_1 := 2\frac{\log n}{n} + \varepsilon_0,$$
$$p_2 := 2\frac{\log n}{n} + 2\varepsilon_0,$$
$$p_3 := 2\frac{\log n}{n} + 3\varepsilon_0\,.$$

Define $V^{(4)}$ to be the set of all 4-tuples of pairwise distinct vertices of $\mathcal{G}$. We say that such a tuple $(w, x, y, z) \in V^{(4)}$ of $\mathcal{G}$ forms a *square* if $\{\lambda(wx), \lambda(yz)\} \subset [p_1, p_2]$ and $\{\lambda(xy), \lambda(wz)\} \subset [p_2, p_3]$. Note that in a square each vertex can reach any other by a temporal path with labels from $[p_1, p_3]$. Our first goal is to show that a.a.s. there exists sufficiently many squares. This will guarantee that at least one of them can be chosen as a pivot 4-cycle.

Let S be the number of squares in $\mathcal{G}$. Clearly $\mathbb{E}[S] = n^{(4)}\varepsilon_0^4$, where $n^{(i)} := n \cdot (n-1) \cdot (n-2) \cdots (n-i+1)$ denotes the falling factorial. To bound the variance of S, we need to investigate

$$\mathbb{E}[S^2] = \sum_{A \in V^{(4)}} \sum_{B \in V^{(4)}} \mathbb{P}[A, B \text{ are both squares}] = \sum_{i=4}^{8} \sum_{\substack{A,B \in V^{(4)} \\ |A \cup B| = i}} \mathbb{P}[A, B \text{ are both squares}]\,.$$

For $i = 8$ (i.e., A, B being vertex-disjoint) we clearly have $n^{(8)}$ summands, whereas for $i \leq 7$ the number of pairs $A, B \in V^{(4)}$ with $|A \cup B| = i$ can be upper-bounded by $(8n)^i \leq 8^7 n^i$. Clearly A and B can have at most $8 - i$ shared edges and thus $\mathbb{P}[A, B \text{ are both squares}] \leq \varepsilon_0^i$. Therefore,

$$\mathbb{E}[S^2] \leq n^{(8)}\varepsilon_0^8 + \sum_{i=4}^{7} 8^7 n^i \varepsilon_0^i \leq \mathbb{E}[S]^2 + 8^7 \cdot \sum_{i=4}^{7} 4^i (\log n)^{0.8i} \leq \mathbb{E}[S]^2 + 8^7 \cdot 4 \cdot 4^7 (\log n)^{5.6}$$

and thus the variance $\mathbb{D}[S]$ is no more than $8^7 \cdot 4^8 (\log n)^{5.6} \in o(\mathbb{E}[S]^2)$. By Chebyshev's inequality, we obtain that $S \geq \mathbb{E}[S]/2$ a.a.s.

Assume now a square (w, x, y, z) to be given and let $\mathcal{G}'$ be the temporal subgraph of $\mathcal{G}$ obtained by deleting the vertices x, y, z. Observe that $\mathcal{G}'$ is distributed as an element of $\mathcal{F}_{n-3,\,1}$ as none of the square's edges are contained within. Note that the intervals $[0, p_1]$ and $[p_3, p]$ both have length $\frac{2\log n}{n} + \varepsilon_0$. Thus, by Theorem 5.2, with probability at least $1 - \frac{10}{\log(n-3)}$, vertex w is a temporal sink of $\mathcal{G}'_{[0,p_1]}$ and also a temporal source in $\mathcal{G}'_{[p_3,p]}$. In this case, we call (w, x, y, z) a *good square*. Then, by taking the square (w, x, y, z) together with the foremost tree from w in $\mathcal{G}'_{[p_3,p]}$ and the "reverse foremost tree" to w in $\mathcal{G}'_{[0,p_1]}$, we obtain a spanner of $\mathcal{G}_{[0,p]}$ that has $4 + 2(n-4) = 2n - 4$ edges.

It only remains to show that there exists a good square. According to the above, the expected number of bad (i.e., non-good) squares S_{bad} satisfies $\mathbb{E}[S_{\text{bad}}] \leq \mathbb{E}[S]\frac{10}{\log(n-3)}$. By Markov's inequality and our bound on S, the probability for every square of $\mathcal{G}$ to be bad is at most

$$\mathbb{P}[S_{\text{bad}} = S] \leq \mathbb{P}\left[S < \frac{\mathbb{E}[S]}{2}\right] + \frac{\mathbb{E}[S_{\text{bad}}]}{(\mathbb{E}S)/2} \leq o(1) + \frac{20}{\log(n-3)} \in o(1)\,.$$

Thus, $\mathcal{G}$ has a good square a.a.s. $\square$

While there is a gap between the lower bound $3\log n/n$ and the upper bound $4\log n/n$, we conjecture that the upper bound is a sharp threshold for existence of an optimal spanner.

Conjecture 5.8. *The sharp threshold for existence of an optimal spanner in a random simple temporal graph exists and is equal to* $4\frac{\log n}{n}$.

6 Application to gossiping and population protocols

In this section, we demonstrate versatility and flexibility of our model of random temporal graphs (RSTGs) by presenting a number of implications that our results give for sequential gossiping and population protocols.

In the classical gossiping setting, there are n agents each of which knows a single secret. The agents can communicate via telephone calls. Whenever an agent calls another agent, the two exchange all secrets they know. An agent who learns all n secrets becomes an *expert*. The standard goal is to achieve the state in which all agents are experts. A sequence of calls leading to this state is called a *gossiping protocol*.

In the theory of population protocols, the same model appears in disguise of an epidemic spreading model, which is usually utilized in the study of various population protocol problems. In this model a *population* consists of n agents. Initially, each agent has a unique piece of information. At every discrete time step a *scheduler* picks a pair of agents to interact. The two interacting agents transmit to each other the pieces of information they have. Depending on a specific problem at hand, a goal of interest might be to achieve the state where every agent has received a specific piece of information or every agent has received all the pieces of information (among other, more sophisticated interaction patterns).

Clearly the two models are equivalent and for the sake of concreteness, in what follows, we mostly stick to the terminology of gossiping protocols. The model can further be specified by imposing additional restriction on the communication model. In particular, we will discuss two most common and natural models (following the notations from [12]):

- `ANY` no restrictions, i.e., a next call can happen between any two agents independently of the previous calls;
- `CO` (stands for *call-once*) no call can be repeated, i.e., there can be at most one call between any two agents.

According to a classical result from the '70s, any gossiping protocol requires at least $2n - 4$ calls (see [7] for a historical note on the corresponding combinatorial problem). Almost at the same time, the duration of gossiping protocols has been studied in a randomized setting in which every next call happens uniformly at random among all eligible calls (in the terminology of population protocols, there is a *probabilistic scheduler* who picks a pair of agents to interact uniformly at random). In a sequence of three papers [19, 8, 14] with the same title "*Random exchanges of information*" asymptotics for the *expected* number of calls until a fixed agent becomes an expert and until all agents become experts were obtained for the randomized `ANY` model of communication, i.e., for the model where every next call happens between a pair of distinct agents chosen uniformly at random among all pairs of distinct agents.

Theorem 6.1 ([19, 8, 14])**.** *In the randomized* `ANY` *model of communication,*

1. *the expected number of calls until a* fixed *agent becomes an expert is* $n \log n + O(n)$*;*
2. *the expected number of calls until* all *agents become experts is* $\frac{3}{2} n \log n + O(n)$*.*

A few year ago, it was shown in [18] that the number of calls until a fixed agent becomes an expert is concentrated around its expected value (with controlled convergence speed).

Theorem 6.2 ([18, 17])**.** *In the randomized* `ANY` *model of communication, a.a.s. the number of calls until a* fixed *agent becomes an expert is* $n \log n \cdot (1 + o(1))$*.*

Even more recently, it was shown in [10] that the number of calls until all agents become experts does not exceed its expected value a.a.s. We remark that the actual lemma in [10] states a weaker bound on time (to achieve a stronger bound on the probability), but an intermediate step of the proof directly implies the following:

Theorem 6.3 ([10], proof of Lemma A.7)**.** *In the randomized* `ANY` *model of communication, a.a.s. the number of calls until* all *agents become experts is* at most $\frac{3}{2}n \log n \cdot (1 + o(1))$ *a.a.s.*

According to [12], there seem to be no similar estimations known for the randomized `CO` model, i.e., for the model where every next call happens between a pair of distinct agents chosen uniformly at random among all pairs agents that did not interact before.

In Section 6.1, we show how our results readily translate in terms of numbers of calls until various events happen in the randomized `CO` model, including the event that a fixed agent becomes an expert and the event that all agents become experts. Quite significantly, we provide asymptotic estimates for the actual number of calls, rather than for its expected value. In Section 6.2, we demonstrate the flexibility of our model by showing how it can be modified to correspond to the randomized `ANY` model. All our results are preserved in the modified model, which thus gives estimates that are similar to the ones in the randomized `CO` model.

On the one hand, we significantly advance old and recent known results, and on the other hand, we provide new results for natural events such as the number of calls until the first expert appears and until two fixed agents exchange their secrets. Our results also imply that information essentially propagates at the same speed in randomized `CO` and randomized `ANY`.

6.1 Random information exchanges in the CO model

Recall that in our model $\mathcal{F}_{n,1}$, the time labels induce edge orderings of the complete graph, each of which is equiprobable. The same distribution of edge orderings is obtained if we construct a random ordering by choosing edges one by one uniformly at random among all edges that are not yet chosen, i.e., in the same way as calls are originated in the randomized `CO` model. Therefore, by interpreting edges as calls and time labels as ranks of the calls, the number of edges in $\mathcal{G}_{[0,p]}$, where $\mathcal{G} \sim \mathcal{F}_{n,1}$ is the number of calls that happened no later than time p. Furthermore, since the underlying graph of $\mathcal{G}_{[0,p]}$ is distributed according to $G_{n,p}$ and the number of edges in $G_{n,p}$ is concentrated around $\binom{n}{2}p$ when $p = \Theta(\log n/n)$, our results about random temporal graphs from Section 5 translate straightforwardly to the following results in the randomized `CO` model.

Theorem 6.4. *In the randomized* `CO` *model, a.a.s.*

1. *the number of calls until two fixed agents exchange their secrets is* $\frac{1}{2}n \log n \cdot (1 + o(1))$*;*
2. *the number of calls until at least one of the agents becomes an expert is* $n \log n \cdot (1 + o(1))$*;*
3. *the number of calls until a fixed agent becomes an expert is* $n \log n \cdot (1 + o(1))$*;*
4. *the number of calls until all agents become experts is* $\frac{3}{2}n \log n \cdot (1 + o(1))$*.*

6.2 Random information exchanges in the ANY model

By replacing the uniform distribution of edge labels with another distribution, we can obtain random temporal graphs with different temporal dynamics. This flexibility, in particular, allows us to simulate the randomized `ANY` model. In order to achieve this, we replace the uniform distribution for edge labels in $\mathcal{F}_{n,1}$ with a suitable Poisson point process.

More specifically, we introduce a random temporal graph model $\mathcal{H}_n$, in which the labels of each of the $\binom{n}{2}$ potential edges appear independently according to a Poisson point process with rate 1 starting at time 0 and running infinitely long. In this model each edge gets a countably infinite set of labels with probability 1. We also define a finite random temporal graph model $\mathcal{H}_{n,p} := (\mathcal{H}_n)_{[0,p]}$ where the process is stopped at time p. Note that, in $\mathcal{H}_{n,p}$, an edge may appear an arbitrary number of times, including zero in which case the edge is not present in the underlying graph. The expected number of appearances for each edge is exactly p. Furthermore, as the Poisson distribution has variance equal to the expectation, by Chebyshev's inequality the number of appearance of a fixed edge is concentrated around its expected value. Hence, the total number of edge appearances in $\mathcal{H}_{n,p}$ is concentrated around $\binom{n}{2}p$.

Since all edges appear according to independent identical Poisson point processes, at any fixed point in time, the next edge to appear is distributed uniformly at random among all $\binom{n}{2}$ possible edges, i.e., in the same way as calls are scheduled in the randomized `ANY` model. Hence, if p_0 is a threshold probability for a temporal property in $\mathcal{H}_{n,p}$, then it translates to a $\binom{n}{2}p_0$-calls threshold in the randomized `ANY` model for the corresponding property.

In this section, we will show that all our main results about $\mathcal{F}_{n,p}$ can be transferred to $\mathcal{H}_{n,p}$ with only minor changes in the proofs. Precisely, the corresponding results from Section 5 translate to the following results in the randomized `ANY` model.

Theorem 6.5. *In the randomized `ANY` model, a.a.s.*

1. *the number of calls until two fixed agents exchange their secrets is* $\frac{1}{2}n\log n \cdot (1+o(1))$*;*
2. *the number of calls until at least one of the agents becomes an expert is* $n\log n \cdot (1+o(1))$*;*
3. *the number of calls until a fixed agent becomes an expert is* $n\log n \cdot (1+o(1))$*;*
4. *the number of calls until all agents become experts is* $\frac{3}{2}n\log n \cdot (1+o(1))$*.*

In the remainder of this section, we explain which changes (if any) need to be made to the statements of Sections 3 to 5 in order to apply to the $\mathcal{H}_{n,p}$ model.

Results based on 2-hop approach (Section 3)

Lemma 6.6 (cf. Lemma 3.1)**.** *Let $\alpha \geq 3$ and let $p = \alpha\sqrt{\log n/n}$. Then, for all $n \geq 4$ and $p \leq 1$, an arbitrary vertex of $(G,\lambda) \sim \mathcal{H}_{n,p}$ is a temporal source with probability at least $1 - n^{-\alpha^2/12+1}$.*

Proof. The construction with intermediate vertices stays the same. The probability $\mathbb{P}[S_{yz}]$ that vertex x can reach y in two steps via y becomes $1-e^{-p}(1+p) > 1-\exp(-p+p-p^2/2+p^3/3) = 1-\exp(-p^2/2+p^3/3) > 1-\exp(-p^2/6)$. Here the first formula is a standard property of the Poisson process, the first inequality uses the Taylor series of the natural logarithm, and the second uses $p \leq 1$. The probability p_1 of all $n-2$ intermediate vertices being unsuitable is at most $(1-(1-\exp(-p^2/6)))^{n-2} = \exp(-(p^2/6)(n-2)) = \exp(-\alpha^2(\log n)(n-2)/(6n)) \leq n^{-\alpha^2/12}$ by independence. The rest of the proof is the same as before with $\alpha^2/12$ instead of $\alpha^2/4$. $\square$

Just like Corollary 3.2, we obtain the following corollary.

Corollary 6.7 (cf. Corollary 3.2)**.** *Let $p = \frac{\log n}{\sqrt{n}}$. Then, $(G,\lambda) \sim \mathcal{H}_{n,p}$ is temporally connected with probability at least $1 - n^{-\frac{\log n}{12}+2}$.*

Results on foremost tree evolution (Section 4)

A change in the foremost tree algorithm (Algorithm 1) is necessary to account for the multiple labels on some edges. The algorithm is now applied to the graphs distributed according to $\mathcal{H}_n$. The change is similar to treating each edge label as a separate edge. More specifically, for each edge e in S_k, i.e., an edge connecting two vertices with exactly one of them included in the tree T_{k-1}, we consider the earliest label $\lambda_{e,k} \in \lambda(e)$ such that $T_{k-1} \cup \{(e, \lambda_{e,k})\}$ is an increasing temporal tree. With probability 1, such a label exists. Just like before, we select the edge e with the minimal possible $\lambda_{e,k}$.

Apart from the fact that (G, λ) is now sampled from $\mathcal{H}_n$ and notation changes to use $\lambda_{e,k}$ instead of $\lambda(e)$ where appropriate, the definitions of T_k^v, Y_k^v, c_k, $\widehat{Y}_k^v$ and $\mathcal{A}_k^v$ apply without changes, as do Lemma 4.1 and Lemma 4.2:

Lemma 6.8 (cf. Lemma 4.1)**.** *Let (G, λ) be a temporal graph (not necessarily simple) and v be a temporal source in (G, λ). Then*

(i) modified Algorithm 1 constructs a foremost tree for v in (G, λ);

(ii) $\lambda_{e_1,1} \leq \lambda_{e_2,2} \leq \ldots \leq \lambda_{e_{n-1},n-1}$.

Proof. Apart from notation changes, this lemma is proven exactly the same as before. □

Lemma 6.9 (Properties of c_k, cf. Lemma 4.2)**.** *We have*

(i) $\sum_{i=1}^{n-1} c_i^2 \leq \frac{64(\log\log n)^2}{n^2}$;

(ii) for a fixed vertex v, with probability at least $1 - 4/\log n$ the equality $\widehat{X}_k^v = X_k^v$ holds for every $k \in [n-1]$.

Proof. Only the proof of (ii) requires a minor adaption. We have that

$$\mathbb{P}\big[\widehat{X}_k^v \neq X_k^v\big] = \mathbb{P}\big[X_k^v > c_k\big] = e^{-k(n-k)c_k}$$

after which we may again follow the proof of Lemma 4.2. □

Instead of Lemma 4.3, we may state a slightly stronger result here which also replaces Corollary 4.4:

Lemma 6.10 (cf. Lemma 4.3)**.** *For every $k \in [n-1]$ we have*

(i)
$$\mathbb{E}[X_k^v \mid \mathcal{A}_{k-1}^v] = \frac{1}{k(n-k)};$$

(ii)
$$\frac{1 - 1/\log n}{k(n-k)} \leq \mathbb{E}[\widehat{X}_k^v \mid \mathcal{A}_{k-1}^v] \leq \frac{1}{k(n-k)}.$$

Proof. (i): Due to the memorylessness of the exponential distribution, X_k^v conditioned on $\mathcal{A}_{k-1}^v$ is distributed as the minimum of $k(n-k)$ independently Exp(1)-distributed random variables. Thus, the distribution of X_k^v is exactly $\mathrm{Exp}(k(n-k))$ and as such it has expected value $\frac{1}{k(n-k)}$.

(ii): The only difference to the proof of Lemma 4.3 is in the evaluation of the integral

$$\begin{aligned}
\int_0^{c_k} \mathbb{P}[X_k^v \geq t \mid A_{k-1}^v]\,\mathrm{d}t &= \int_0^{c_k} \mathbb{P}[X_k^v \geq t]\,\mathrm{d}t \\
&= \int_0^{c_k} \exp\left(-k(n-k)t\right)\mathrm{d}t \\
&= \frac{1 - \exp(-k(n-k)c_k)}{k(n-k)}.
\end{aligned}$$

Then, the desired bound follows from the fact that $\exp(-k(n-k)c_k) \leq \frac{1}{\log n}$ (see Eq. (1)). □

The statements of Lemma 4.6 and Theorem 4.7 apply unchanged, with their proofs only requiring trivial modifications. Similarly, Corollary 4.8 does not rely on the specific details of the distributions.

Results on sharp thresholds for temporal graph properties (Section 5)

Theorems 5.1 to 5.3 (sharp thresholds for point-to-point reachability, temporal source, and first temporal source) apply without modifications as they follow solely from the concentration results obtained in Section 4.2. Lemma 5.5 only requires a minor change to the proof:

Lemma 6.11 (cf. Lemma 5.5)**.** *Let* $p < \frac{3 \log n}{n} - \varepsilon(n)$*, where* $\varepsilon(n) = \frac{6(\log n)^{0.8}}{n}$*, and let* $\mathcal{G} \sim \mathcal{H}_n$*. Then, a.a.s. the temporal graph* $\mathcal{G}_{[0,p]}$ *contains at least one vertex which is not a temporal sink.*

Proof. The proof works the same, except that the probability for a pair of vertices to form an edge in the underlying graph H of $\mathcal{G}_{[q,p]}$ is always

$$u = 1 - e^{q-p} \le p - q \le \frac{3 \log n}{n} - \varepsilon - \frac{2 \log n}{n} + \frac{\varepsilon}{2} < \frac{\log n - (\log n)^{0.8}}{n},$$

In particular, H is distributed as an element of $G_{n,u}$ and the remainder of the proof may be copied as is. ☐

The proof of Lemma 5.6 is simplified due to the fact that we do not need to bound $\Delta(\mathcal{G}_{[0,q]})$ anymore.

Lemma 6.12 (cf. Lemma 5.6)**.** *Let* $p > \frac{3 \log n}{n} + \varepsilon(n)$*, where* $\varepsilon(n) = \frac{3(\log n)^{0.8}}{n}$*, and let* $\mathcal{G} \sim \mathcal{H}_n$*. Then, a.a.s. every vertex in* $\mathcal{G}_{[0,p]}$ *is a temporal sink.*

Proof. We define q and r and S as in Lemma 5.6, and observe that $|V \setminus S| \le r$ a.a.s. The set C_w of potential edges connecting $w \in V \setminus S$ to an element of S is simplified to

$$C_w = \{vw \mid v \in S\}$$

which now trivially implies $|C_w| = n - 1 - r \ge n - 2r = d$. As the waiting time of each edge is exponentially distributed, we now get $\mathbb{P}[T_w > x] \le e^{-xd}$. In particular, the estimation

$$\tau := \mathbb{P}\left[T_w > \frac{\log n}{n}\right] \le e^{-(\log n - 2 \log \log \log n)}$$

still applies, so the remainder of the proof may be copied without modifications. ☐

Thus, we again obtain Theorem 5.4, i.e., the sharp threshold on temporal connectivity in $\mathcal{H}_{n,p}$. Regarding optimal spanners, Theorem 5.7 remains to be checked.

Theorem 6.13 (cf. Theorem 5.7)**.** $\mathcal{H}_{n,p}$ *a.a.s. has an optimal spanner if* $p \ge 4\frac{\log n}{n} + \varepsilon(n)$ *where* $\varepsilon(n) := \frac{16(\log n)^{0.8}}{n} \in o\left(\frac{\log n}{n}\right)$.

Proof. In comparison to the proof of Theorem 5.7, only the probability of an edge appearing in an interval of length ε_0 changes from ε_0 to $1 - e^{-\varepsilon_0}$. Thus, $\mathbb{E}[S] = n^{(4)}(1 - e^{-\varepsilon_0})^4$ and

$$\begin{aligned}
\mathbb{E}[S^2] &\le n^{(8)}(1 - e^{-\varepsilon_0})^8 + \sum_{i=4}^{7} 8^7 n^i (1 - e^{-\varepsilon_0})^i \\
&\le \mathbb{E}[S]^2 + 4 \cdot 8^7 n^7 (1 - e^{-\varepsilon_0})^7 \\
&\le \mathbb{E}[S]^2 + 4 \cdot 8^7 n^7 \varepsilon_0^7
\end{aligned}$$

where the second inequality uses the fact that $1 - e^{-\varepsilon_0} \ge \varepsilon_0 - \varepsilon_0^2/2 \ge \varepsilon_0/2 \ge 1/n$. Therefore, the bound on the variance $\mathbb{D}[S]$ and the remainder of the proof remain in place. ☐

7 Conclusion

In this paper, we have presented a natural model of random temporal graphs called RSTG, in which every edge of an Erdős–Rényi graph $G_{n,p}$ is assigned a single presence time chosen uniformly at random in the unit interval $[0, 1]$. The study of various forms of temporal reachability in these graphs revealed a rich diversity of thresholds, in stark contrast with static graphs. Put together, these thresholds offer a measure of the discrepancy between static and temporal connectivity in random graphs. Despite the simplicity of the model, we have shown that RSTGs capture, in a scale-preserving way, several classical phenomena observed in gossip theory and population protocols. Furthermore, some of our results in RSTGs were shown to strengthen and/or to complete existing results in these fields. All of our characterizations but one correspond to sharp threshold. The last one concerns the existence of optimal temporal spanners (i.e., spanners of size $2n - 4$) in RSTGs, for which we prove an upper bound of $p = 4 \log n/n$. Whether this bound is actually a sharp threshold is left open; we conjecture that it is (Conjecture 5.8). Now that the most basic phase transitions in temporal reachability are characterized in RSTGs, it would be quite natural to start looking at more complex properties, e.g. motivated by applications in networking. For example, how long does it take before back-and-forth temporal paths exist between all pairs of vertices (i.e., information dissemination with acknowledgment)? Can one establish sharp thresholds for existence of temporal spanners of some larger given size? We hope that the simple model we considered in this paper, together with the results we presented on basic forms of reachability, have now paved the way to further investigations of more complex phenomena in random temporal graphs.

References

[1] Omer Angel, Asaf Ferber, Benny Sudakov, and Vincent Tassion. "Long Monotone Trails in Random Edge-Labellings of Random Graphs". In: *Combinatorics, Probability and Computing* 29.1 (2020), pp. 22–30. DOI: 10.1017/S096354831900018X.

[2] Dana Angluin, James Aspnes, Zoë Diamadi, Michael J. Fischer, and René Peralta. "Computation in networks of passively mobile finite-state sensors". In: *Proceedings of the 23th Annual ACM Symposium on Principles of Distributed Computing (PODC)*. Ed. by Soma Chaudhuri and Shay Kutten. ACM, 2004, pp. 290–299. DOI: 10.1145/1011767.1011810.

[3] Kyriakos Axiotis and Dimitris Fotakis. "On the Size and the Approximability of Minimum Temporally Connected Subgraphs". In: *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*. 2016, 149:1–149:14. DOI: 10.4230/LIPIcs.ICALP.2016.149.

[4] Kazuoki Azuma. "Weighted sums of certain dependent random variables". In: *Tohoku Mathematical Journal* 19.3 (1967), pp. 357–367. DOI: 10.2748/tmj/1178243286.

[5] Sandeep Bhadra and Afonso Ferreira. "Complexity of connected components in evolving graphs and the computation of multicast trees in dynamic networks". In: *Proceedings of the 2nd International Conference on Ad-Hoc Networks and Wireless (ADHOC-NOW)*. 2003, pp. 259–270. DOI: 10.1007/978-3-540-39611-6_23.

[6] Béla Bollobás. *Random Graphs*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511814068.

[7] Béla Bollobás. *The art of mathematics: Coffee time in Memphis*. Cambridge University Press, 2006.

[8] David W Boyd and J Michael Steele. “Random exchanges of information”. In: *Journal of Applied Probability* 16.3 (1979), pp. 657–661. DOI: 10.2307/3213094.

[9] Richard T. Bumby. “A Problem with Telephones”. In: *SIAM Journal on Algebraic Discrete Methods* 2.1 (1979). DOI: 10.1137/0602002.

[10] Janna Burman, David Doty, Thomas Nowak, Eric E. Severson, and Chuan Xu. “Efficient self-stabilizing leader election in population protocols”. In: *CoRR* abs/1907.06068 (2019). arXiv: 1907.06068.

[11] Arnaud Casteigts, Joseph Peters, and Jason Schoeters. “Temporal Cliques Admit Sparse Spanners”. In: *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*. Vol. 132. 2019, 134:1–134:14. DOI: 10.4230/LIPIcs.ICALP.2019.134.

[12] Hans van Ditmarsch, Ioannis Kokkinis, and Anders Stockmarr. “Reachability and expectation in gossiping”. In: *Proceedings of PRIMA 2017: Principles and Practice of Multi-Agent Systems*. Springer. 2017, pp. 93–109. DOI: 10.1007/978-3-319-69131-2_6.

[13] Alan Frieze and Michał Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015. DOI: 10.1017/CBO9781316339831.

[14] John Haigh. “Random exchanges of information”. In: *Journal of Applied Probability* 18.3 (1981), pp. 743–746. DOI: 10.2307/3213330.

[15] David Kempe, Jon Kleinberg, and Amit Kumar. “Connectivity and inference problems for temporal networks”. In: *Proceedings of the 32nd annual ACM symposium on Theory of computing (STOC)*. 2000, pp. 504–513. DOI: 10.1145/335305.335364.

[16] Mikhail Lavrov and Po-Shen Loh. “Increasing Hamiltonian paths in random edge orderings”. In: *Random Structures & Algorithms* 48.3 (2016), pp. 588–611. DOI: 10.1002/rsa.20592.

[17] Yves Mocquard, Bruno Sericola, and Emmanuelle Anceaume. “Probabilistic Analysis of Rumor-Spreading Time”. In: *Informs Journal on Computing* 32.1 (2020), pp. 172–181. DOI: 10.1287/ijoc.2018.0845.

[18] Yves Mocquard, Bruno Sericola, Samantha Robert, and Emmanuelle Anceaume. “Analysis of the propagation time of a rumour in large-scale distributed systems”. In: *Proceedings of the 15th IEEE International Symposium on Network Computing and Applications*. Ed. by Alessandro Pellegrini, Aris Gkoulalas-Divanis, Pierangelo di Sanzo, and Dimiter R. Avresky. IEEE Computer Society, 2016, pp. 264–271. DOI: 10.1109/NCA.2016.7778629.

[19] JW Moon. “Random exchanges of information”. In: *Nieuw Archief voor Wiskunde* 20 (1972), pp. 246–249.