Generating Temporal Contact Graphs Using Random Walkers

Almasan, Anton David; Shvydun, Sergey; Scholtes, Ingo; Mieghem, Piet Van

DOI
10.1109/TNSE.2025.3537162

Publication date
2025

Document Version
Final published version

Published in
IEEE Transactions on Network Science and Engineering

Generating Temporal Contact Graphs Using Random Walkers

Anton-David Almasan, Sergey Shvydun, Ingo Scholtes, and Piet Van Mieghem, *Fellow, IEEE*

***Abstract*—We study human mobility networks through time-series of contacts between individuals. Our proposed Random Walkers Induced temporal Graph (RWIG) model generates temporal graph sequences based on independent random walkers that traverse an underlying graph in discrete time steps. Co-location of walkers at a given node and time defines an individual-level contact. RWIG is shown to be a realistic model for temporal human contact graphs, which may place RWIG on a same footing as the Erdos–Renyi (ER) and Barabasi–Albert (BA) models for fixed graphs. Moreover, RWIG is analytically feasible: we derive closed form solutions for the probability distribution of contact graphs.**

***Index Terms*—Temporal networks, generative models, network dynamics, Markov process, random walks.**

I. Introduction

In THE past years, the study of temporal graphs has received a surge of interest, e.g. to model how time-varying human contact patterns impact epidemics like COVID-19 [1], [2], [3]. Empirical studies of real-world contact patterns have identified several characteristics of temporal graphs that can influence dynamical processes.

A first line of research has focused on the question how the *temporal distribution of interactions* affects the evolution of dynamical processes in temporal graphs. Studies on the influence of non-Poissonian and bursty node activity patterns [4], [5], [6], [7] or long-lasting or concurrent interactions [8], [9] have shown that real contact patterns can both slow down or speed up spreading processes compared to a static graph, where all links are always active.

A second line of research has addressed the question *how the temporal ordering of interactions* influences dynamical processes such as diffusion or epidemic spreading. In a nutshell, for two temporal contacts occurring between Alice and Bob at time t and between Bob and Carol at time t', a virus can only spread from Alice via Bob to Carol if the contact between Alice and Bob occurs *before* the contact between Bob and Carol, i.e. iff $t<t'$. If the temporal ordering of contacts is reversed, no *time-respecting path* exists between Alice and Carol due to the directedness of the *arrow of time*. Empirical studies on social, biological, and technical systems [10], [11], [12] have shown that the *causal topology* of temporal graphs, i.e. who can influence whom via time-respecting paths, is more complex than what we expect from their static, time-aggregated counterparts, leading to non-trivial effects such as a speed up or slow down of diffusion processes compared to (randomized) temporal graphs, that lack correlations in the temporal ordering of interactions [13], [14], [15], [16], [17], [18].

Several temporal graph modelling and learning approaches have been proposed that account for some of the complex characteristics of empirical contact patterns [19], [20], [21], [22], [23], [24]. Recently, a system theoretical approach towards emulating temporal graphs is presented in [25]. Various approaches to model human mobility, which could explain some of the temporal characteristics of human contact patterns, are discussed in [26], [27], [28], [29]. Beside epidemic spreading, a better understanding of temporal mechanisms can also facilitate the design, management and control of mobile opportunistic networks [30] or human mobility in public transportation networks [31]. However, we still lack simple *generative models for temporal graphs* that (i) are able to reproduce realistic contact patterns, (ii) facilitate analytic treatment and (iii) shed light on potential mechanisms that shape both the topological and temporal dimension of temporal graphs.

Addressing this research gap, we propose the *Random Walkers Induced temporal Graph* (*RWIG*) model, which uses multiple random walkers on a finite graph as a generative model for temporal contact networks. Any realization of a discrete-time Markov process on N states can be represented by a random walk on the corresponding Markov graph with N nodes (states), where a link between two states i and j is characterised, i.e. both directed and weighted, by the transition probability p_{ij}. The RWIG model considers a collection of M random walkers that simultaneously traverse the Markov graph in discrete timesteps according to the $N\times N$ Markov transition probability matrix P with elements equal to the transition probabilities p_{ij}. Hence, each walker executes a realization of the same Markov process or, equivalently, each walker's trajectory is driven by

Received 11 July 2024; revised 9 January 2025; accepted 23 January 2025. Date of publication 30 January 2025; date of current version 25 April 2025. This work was supported by European Research Council (ERC) through European Union's Horizon 2020 Research and Innovation Program under Grant 101019718. The work of Ingo Scholtes was supported by Swiss National Science Foundation under Grant 176938. Recommended for acceptance by Dr. Dusit Niyato. *(Corresponding author: Anton-David Almasan.)*

Anton-David Almasan, Sergey Shvydun, and Piet Van Mieghem are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: A.D.Almasan@tudelft.nl; S.Shvydun@tudelft.nl; P.F.A.VanMieghem@tudelft.nl).

Ingo Scholtes is with the Chair of Machine Learning for Complex Networks, Center for Artificial Intelligence and Data Science (CAIDAS), Julius-Maximilians-Universität Würzburg, 50 D-97074 Würzburg, Germany (e-mail: ingo.scholtes@uni-wuerzburg.de).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSE.2025.3537162, provided by the authors.

Digital Object Identifier 10.1109/TNSE.2025.3537162

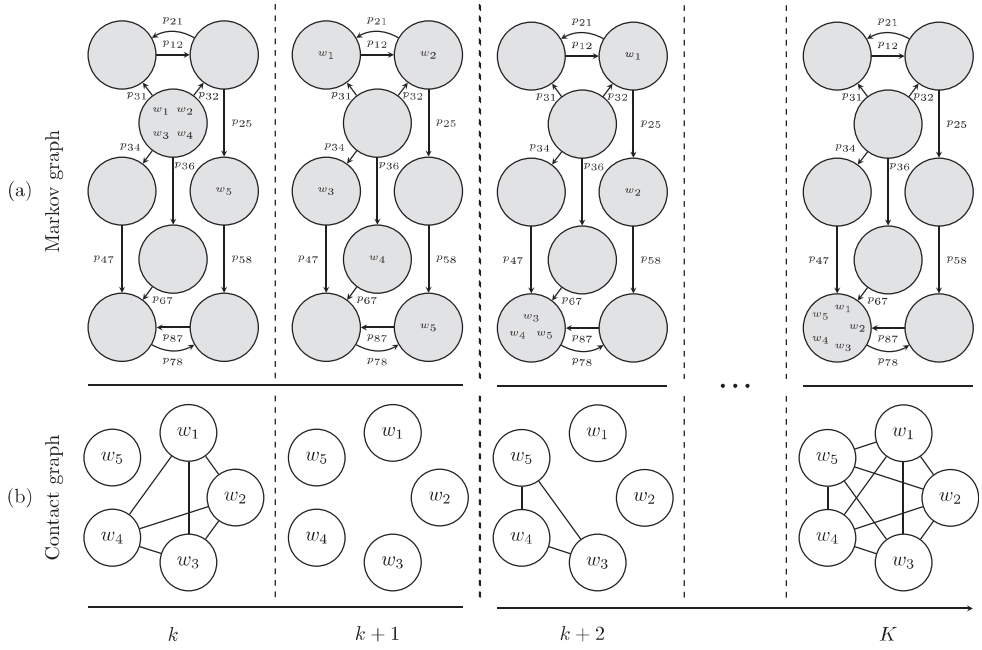

Fig. 1. Contact graph generation using RWIG: (a) A collection $\mathcal{M} = \{w_1, w_2, w_3, w_4, w_5\}$ of random walkers simultaneously traverse the Markov graph. The Markov states (shaded) are nodes and the directed links are the transition probabilities. (b) At each discrete time step $k \in \{1, \ldots, K\}$, the contact graph G_k between walkers is formed. RWIG generates a link in the contact graph G_k between each pair of walkers found in the same Markov state at discrete time k.

the Markov process. Thus, we assume in RWIG that the Markov process generates human mobility trajectories over a set of places (states). Next to the Markov graph, at discrete time k, the contact graph G_k with M nodes is generated, in which the nodes represent the random walkers. The main assumption of RWIG is that *links in the contact graph G_k are created between walkers which visit the same state in the Markov graph at discrete time k*. Fig. 1 exemplifies $M = 5$ random walkers, who traverse a Markov graph with $N = 8$ states (shaded) in discrete-time steps according to the transition probabilities p_{ij}. The probabilities p_{ij} are depicted in Fig. 1(a) on the links between Markov states. The observation window has length K as displayed on the horizontal axis and four discrete time steps are shown. In Fig. 1(b), RWIG generates the contact graph of the 5 walkers at each timestep by creating links between all walkers found in the same state in the Markov graph. For instance, at time $k + 2$ walkers w_3, w_4, w_5, which were in different states at time $k + 1$, move to the same state and, consequently, form a fully connected subgraph (or clique) of size 3. Conversely, walkers w_1, w_2 move to different Markov states and thus become isolated nodes in the contact graph G_{k+2}.

A physical interpretation of RWIG is a collection of individuals moving through space. The underlying graph with adjacency matrix A represents a city map, with nodes as various locations (e.g. restaurants, workplaces, homes, public transport stations, etc) and links as physical paths between locations. The random walkers represent individuals and the transition probabilities p_{ij} assume that all individuals behave the same.

We can regard the probabilities p_{ij}, which together form an $N \times N$ transition probability matrix P, as a common *policy*, which all individuals follow. The transition probability matrix P can generally take the form of any function $f(A)$ of the adjacency matrix A which results in a stochastic matrix [35]. An example in which the probabilities of jumping from a state i to any other adjacent state j are all equal is $P = \Delta^{-1}A$, where Δ is the diagonal matrix of the degree vector of the underlying graph with adjacency matrix A.

As a common policy is restrictive and often unrealistic (e.g. a kindergartener would visit different locations than an office worker), we consider that each random walker w_r can have a different policy, or transition probability matrix P_r. All policies, however, still reflect the same underlying graph topology (e.g. city map). Consequently, if there is no link between two states i and j in the adjacency matrix A, then all policies must have a zero probability for state transitions between nodes i and j (i.e. $a_{ij} = 0$ implies that $(P_r)_{ij} = 0$, for all integers $i, j \in \{1, \ldots, N\}$).

Although the properties of random walks have been extensively studied, the dynamics of multiple random walks on a graph still represents an active research area. Riascos and Sanders [32] study multiple non-interactive random walkers on a graph and analyse the mean encounter times of walkers. A similar model is proposed to generate contacts between individuals in [33], which are then used to study the evolution of epidemics. Masuda et al. [34] present a detailed study of the theory and applications of random walks. To the best of our knowledge, RWIG is the first model which leverages multiple random walks to generate temporal graphs. Our contribution can be summarised:

- We propose the RWIG model based on random walkers for generating temporal contact networks.
- We provide an analytical formula for the probability distribution of the contact graphs, which are produced by RWIG given the transition matrices $\{P_r\}_{r=1}^{M}$ and the initial states of all walkers.

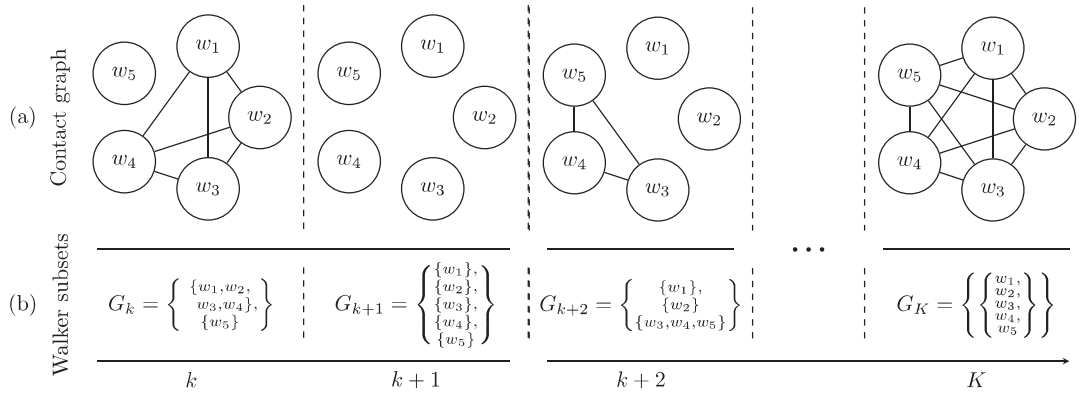

Fig. 2. Contact graphs as partitions: (a) Each contact graph is formed by the union of disconnected cliques (i.e. fully connected graphs), where an isolated node is considered as a clique of size 1. (b) The notion of a clique is abstracted to a subset of pairwise connected walkers. Walkers that belong to different cliques represent elements of disjoint subsets. Each contact graph is represented by disjoint subsets of the walker set $\mathcal{M} = \{w_1, w_2, w_3, w_4, w_5\}$, i.e. a partition where the cells (i.e. subsets) of the partition indicate the cliques.

- We demonstrate how RWIG is able to generate contact graphs that resemble real temporal networks.

The paper is organised as follows. In Section II, we describe the state space and topological structure of contact graphs. Section III provides an analytical formula for the probability distribution of the contact graph formed by a set of walkers, conditioned on the walkers' initial states and policies. Section IV discusses RWIG in the steady state. To motivate the applicability of RWIG, Section V offers simulation results illustrating the wide variety of contact graphs produced by RWIG and compares the RWIG generated sequences with empirical data. Finally, we introduce the notation to the reader in Appendix A and mathematical definitions are deferred to Appendix B.

II. Random Walkers Induced Temporal Graph (RWIG)

A. Formulation of RWIG

We consider an undirected unweighted graph with N nodes and L links that is represented by an $N \times N$ adjacency matrix A, which is the underlying graph. A Markov graph that emulates a random walk on that graph has the $N \times N$ probability transition matrix P. For instance, the transition probability matrix $P = \Delta^{-1} A$, where $\Delta = (d_1, d_2, \ldots, d_N)$ and d_i is the degree of node i, describes a Markov graph [35, p. 108-110] in which there is an equal probability to reach neighbouring states. On that Markov graph, M random walkers, independently of each other, jump from one state to another state per discrete time k, starting from $k = 0$ until some finite time $k = K$, according to the $N \times N$ probability transition matrix P. The trajectory of each random walker $w_j \in \{w_1, \ldots, w_M\}$ across the states of the Markov graph can be regarded as one realization of the Markov process [36], that starts in the state described by the $1 \times N$ vector $s_j[0]$.

B. State Space of RWIG

The fundamental assumption of RWIG is that any pair of walkers that meets at time k in the same Markov state is connected in the contact graph G_k. In other words, if q walkers reside in the same state in the Markov graph at discrete time k, they form a fully connected subgraph, i.e. clique of size q in the contact graph G_k. Consequently, the graph G_k consists of the union of disconnected cliques and G_k is only connected and equal to a complete graph K_M if all M walkers meet in the same state. The induced structure describes the contact graph through pairwise disjoint subsets of walkers, which is exemplified in Fig. 2.

The union of the walker subsets in the node set of contact graph G_k equals the complete walker set $\mathcal{M} = \{w_1, w_2, \ldots, w_M\}$. Since the subsets are pairwise disjoint, each possible contact graph generated by RWIG is equivalent to a partition on the walker set $\mathcal{M}$, whose number of cells is equal to the number of disconnected cliques. Thus, we refer to partitions on the walker set $\mathcal{M}$ and contact graphs interchangeably. Additionally, we also refer to m-partitions on the walker set $\mathcal{M}$ and m-clique contact graphs (i.e. a contact graph with m cliques) interchangeably.

To count the number of possible contact graphs, consider an m-clique contact graph G_k at some time k, which is equivalent to an m-partition π_m on the walker set $\mathcal{M}$. In the contact graph, M walkers occupy m different states, where $m \leq M$. Additionally, the number of occupied states m is upper bounded by the total number of states N in the Markov graph. Therefore, the number of states occupied by walkers m is upper bounded by $\min(N, M)$. The total number of contact graphs $|G_k|$ is obtained by summing the number of all possible m-partitions

$$|G_k| = \sum_{m=0}^{\min(N,M)} \mathcal{S}_M^{(m)}, \tag{1}$$

where $\mathcal{S}_M^{(m)}$ are the *Stirling numbers of the second kind* [37].

If the number of walkers does not exceed the number of walker states (i.e. $M \leq N$), m is upper bounded by $\min(N, M) = M$. Therefore, the total number of partitions on the walker set $\mathcal{M}$ and consequently, the total number of contact graphs is $|G_k| = \sum_{m=0}^{M} \mathcal{S}_M^{(m)} = \mathcal{B}_M$, where $\mathcal{B}_M$ is the *M-th Bell number*. The Bell numbers are explained in Appendix B. Table I illustrates a few examples of the number of contact graphs for various combinations of walker count M and number of Markov states

TABLE I
EXAMPLES OF CONTACT GRAPH STATE SPACE CARDINALITY WITH RESPECT TO M WALKERS AND N STATES

M \ N	5	6	7	8	9	10
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	5	5	5	5	5	5
4	15	15	15	15	15	15
5	52	52	52	52	52	52
6	202	203	203	203	203	203
7	855	876	877	877	877	877
8	3845	4111	4139	4140	4140	4140
9	18002	20648	21110	21146	21147	21147
10	86472	109299	115179	115929	115974	115975

N, where the regime $M \leq N$ is shaded. For instance, if $M = 5$ and $N = 3$, the total number of contact graphs is 41 as there are 25 ways for 5 walkers to occupy 3 states and form a 3-clique graph, 15 ways to occupy 2 states (G_k is a 2-clique graph) and only 1 way to be in the same state (G_k is a complete graph).

Therefore, if the number of walkers does not exceed the number of states $M \leq N$, the number of contact graphs formed by M random walkers is equal to the Bell number $\mathcal{B}_M$. Otherwise, if $M > N$, then we omit m-partitions on the walker set $\mathcal{M}$ where $m > N$ because walkers cannot be found in more cliques than there are Markov states.

C. Contact Graph Probability: Examples

After the enumeration of contact graphs in Section II-B, we now seek to find the probability distribution of the contact graphs G_k, conditioned on the initial state vector $s_j[0]$ of walker w_j and Markov transition matrices P_j for each walker $w_j \in \{w_1, w_2, \ldots, w_M\}$.

A contact graph realisation with m cliques is denoted as $g = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\}$, where $\mathcal{A}_i$ for all $i \in \{1, 2, \ldots, m\}$, represent the cliques formed at a discrete time step. Due to the equivalence between contact graphs and partitions on the walker set $\mathcal{M}$ shown in Fig. 2, the cliques $\mathcal{A}_i$ are functionally subsets of walkers found to be in the same state in the Markov graph at a given time.

We also introduce the set of initial conditions for all walkers: $\mathbf{s}_{\mathcal{M}}[0] = \{s_j[0]\}_{j=1}^{M}$, as well as the set of $N \times N$ transition probability matrices for all walkers $\mathbf{P}_{\mathcal{M}} = \{P_j\}_{j=1}^{M}$.

1) Introductory Example: The simplest contact graph example is the complete graph $G_k = \{\mathcal{M}\}$, where all walkers are found in the same Markov state at discrete time k.

The random variable $X_j[k]$ denotes the state in the Markov graph of walker w_j at discrete time k and $\Pr[X_j[k] = i]$ is the probability that walker w_j is in state i in the Markov graph at discrete time k. The i-th element of the probability state vector $s_j[k]$ for walker w_j at time k is then $(s_j[k])_i = \Pr[X_j[k] = i]$. Only if all M walkers are in the same state at discrete time k, a complete graph K_M is formed. The probability that all M walkers are in state i equals $\prod_{j=1}^{M} \Pr[X_j[k] = i]$, because all random walkers move independently of each other in the Markov graph. Summing the probabilities that all walkers are in state i over all states $i \in \{1, 2, \ldots, N\}$ results in the probability that a complete graph $G_k = \{\mathcal{M}\} \equiv K_M$ is created at discrete time k:

$$\Pr[G_k = \{\mathcal{M}\}] = \sum_{i=1}^{N} \prod_{j=1}^{M} \Pr[X_j[k] = i] = \sum_{i=1}^{N} \prod_{j=1}^{M} (s_j[k]) e_i^T,$$

where e_i is the all-zero row vector with 1 at i-th position [36]. Introducing the Hadamard product [35] of the walkers' state probability vectors $s_1[k] \circ \ldots \circ s_M[k] = \bigodot_{j=1}^{M} s_j[k]$:

$$\Pr[G_k = \{\mathcal{M}\}] = \sum_{i=1}^{N} \prod_{j=1}^{M} (s_j[k]) e_i^T = \sum_{i=1}^{N} \left(\bigodot_{j=1}^{M} s_j[k] \right) e_i^T.$$

Finally, introducing the all-ones vector $u = [1, \ldots, 1]$ yields:

$$\Pr[G_k = \{\mathcal{M}\}] = \left(\bigodot_{j=1}^{M} s_j[k] \right) u^T = \bigodot_{j=1}^{M} \left(s_j[0] P_j^k \right) u^T, \tag{2}$$

where we have used the k-step Markov transition probability formula [36]: $s_j[k] = s_j[0] P_j^k$.

Equation (2) expresses the probability that the M walkers are in the same state in the Markov graph at discrete time k. Since $\mathcal{M}$ is just an example of any M size walker set, (2) is also directly applicable to any walker subset $\mathcal{A}_i \subseteq \mathcal{M}$, where $i \in \{1, 2, \ldots, m\}$. We thus define $\sigma_{\mathcal{A}_i}[k]$ as the probability that walkers of a subset $\mathcal{A}_i \subseteq \mathcal{M}$ are in the same state at discrete time k:

$$\sigma_{\mathcal{A}_i}[k] = \bigodot_{w_j \in \mathcal{A}_i} \left(s_j[0] P_j^k \right) u^T. \tag{3}$$

The implementation of (3) is provided in Appendix C (Algorithm 1). The definition of $\sigma_{\mathcal{A}_i}[k]$ in (3) constitutes the basis of our further analysis, because (3) forms a compact and analytically tractable formula relating contact graph probabilities to the transition probability matrices and initial conditions.

Equation (2) calculates the probability that all walkers are in the same state, which is equivalent to the probability of the 1-clique contact graph or the complete graph K_M. To offer insight into the probability of contact graphs with more than one clique, we first calculate the probabilities of the 2-clique and 3-clique contact graphs. We then state and prove in Section III our main theorem for the probability of a general m-clique contact graph.

2) 2-Clique Contact Graph: Let g be a 2-clique contact graph realisation $g = \{\mathcal{A}_1, \mathcal{A}_2\}$. We consider that the walkers in cliques $\mathcal{A}_1$ and $\mathcal{A}_2$ are in Markov states i and j respectively. Summing over all states i, j where $i \neq j$, the probability $\Pr[G_k = g]$ is:

$$\begin{aligned} \Pr[G_k = g] &= \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\prod_{w_u \in \mathcal{A}_1} s_u[k] e_i^T \right) \left(\prod_{w_v \in \mathcal{A}_2} s_v[k] e_j^T \right) \\ &= \sum_{i=1}^{N} \left(\prod_{w_u \in \mathcal{A}_1} s_u[k] e_i^T \right) \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\prod_{w_v \in \mathcal{A}_2} s_v[k] e_j^T \right). \end{aligned} \tag{4}$$

We rewrite the second sum-product term as:

$$\sum_{\substack{j=1\\ j\neq i}}^{N}\prod_{w_v\in\mathcal{A}_2} s_v[k]e_j^T=\sum_{j=1}^{N}\left(\prod_{w_v\in\mathcal{A}_2} s_v[k]e_j^T\right)-\prod_{w_v\in\mathcal{A}_2} s_v[k]e_i^T.$$

Introducing the definition of $\sigma_{\mathcal{A}}[k]$ in (3) yields:

$$\sum_{\substack{j=1\\ j\neq i}}^{N}\prod_{w_v\in\mathcal{A}_2} s_v[k]e_j^T=\bigodot_{w_v\in\mathcal{A}_2}\left(s_v[0]P_v^k\right)u^T-\prod_{w_v\in\mathcal{A}_2} s_v[k]e_i^T$$

$$=\sigma_{\mathcal{A}_2}[k]-\prod_{w_v\in\mathcal{A}_2} s_v[k]e_i^T. \tag{5}$$

Substituting (5) into (4):

$$\Pr[G_k=g]=\sigma_{\mathcal{A}_2}[k]\sum_{i=1}^{N}\left(\prod_{w_u\in\mathcal{A}_1} s_u[k]e_i^T\right)$$
$$-\sum_{i=1}^{N}\left(\prod_{w_u\in\mathcal{A}_1} s_u[k]e_i^T\prod_{w_v\in\mathcal{A}_2} s_v[k]e_i^T\right).$$

Since $\mathcal{A}_1$ and $\mathcal{A}_2$ are complements w.r.t. the walker set $\mathcal{M}$, then $\mathcal{A}_1\cup\mathcal{A}_2=\mathcal{M}$ and thus:

$$\prod_{w_u\in\mathcal{A}_1} s_u[k]e_i^T\prod_{w_v\in\mathcal{A}_2} s_v[k]e_i^T=\prod_{w_u\in\mathcal{M}} s_u[k]e_i^T.$$

Finally, by (2) and (3):

$$\Pr[G_k=g]=\sigma_{\mathcal{A}_2}[k]\sum_{i=1}^{N}\prod_{w_u\in\mathcal{A}_1} s_u[k]e_i^T-\sum_{i=1}^{N}\prod_{w_u\in\mathcal{M}} s_u[k]e_i^T$$
$$=\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]-\sigma_{\mathcal{M}}[k]. \tag{6}$$

The intuition behind (6) is the inclusion-exclusion principle [36, p. 10-12], where the probability of 2 cliques is equal to the probability that walkers from the subsets $\mathcal{A}_1$ and $\mathcal{A}_2$ are each found in the same states minus (hence, excluding) the probability that all walkers are in the same state.

3) 3-Clique Contact Graph: Let g be a 3-clique contact graph realisation $g=\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3\}$. We consider that the walkers in clique $\mathcal{A}_1$ are in Markov state i_1, the walkers in clique $\mathcal{A}_2$ are in Markov state i_2 and that the walkers in clique $\mathcal{A}_3$ are in Markov state i_3. Summing over all states i_1,i_2,i_3 where $i_1\neq i_2\neq i_3$, the probability of the realisation g is:

$$\Pr[G_k=g]=\sum_{i_1=1}^{N}\sum_{\substack{i_2=1\\ i_2\notin\{i_1\}}}^{N}\sum_{\substack{i_3=1\\ i_3\notin\{i_1,i_2\}}}^{N}\prod_{j=1}^{3}\prod_{w_u\in\mathcal{A}_j} s_u[k]e_{ij}^T. \tag{7}$$

Expanding (7) is possible by observing that the probability $\Pr[G_k=g]$ is equal to the product of clique probabilities minus the probability of *any contact graph obtained by amassing cliques* (e.g. $\mathcal{A}_1\cup\mathcal{A}_2$ or $\mathcal{A}_1\cup\mathcal{A}_2\cup\mathcal{A}_3$). In other words, the event that walkers from each of the cliques $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$ are found in the same state, which has probability $\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]\sigma_{\mathcal{A}_3}[k]$, encompasses the 3 events: walkers occupy the same state, walkers occupy two different states, walkers occupy three different states. The probability of a contact graph with three cliques $g=\{\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3\}$ is

$$\Pr[G_k=g]=\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]\sigma_{\mathcal{A}_3}[k]-\Pr[G_k=\{\mathcal{M}\}]$$
$$-\Pr[G_k=\{\mathcal{A}_1\cup\mathcal{A}_2,\mathcal{A}_3\}]$$
$$-\Pr[G_k=\{\mathcal{A}_1\cup\mathcal{A}_3,\mathcal{A}_2\}]$$
$$-\Pr[G_k=\{\mathcal{A}_2\cup\mathcal{A}_3,\mathcal{A}_1\}].$$

Denoting amassed cliques as $\mathcal{A}_i\cup\mathcal{A}_j=\mathcal{A}_{ij}$:

$$\Pr[G_k=g]=\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]\sigma_{\mathcal{A}_3}[k]-\Pr[G_k=\{\mathcal{M}\}]$$
$$-\Pr[G_k=\{\mathcal{A}_{12},\mathcal{A}_3\}]$$
$$-\Pr[G_k=\{\mathcal{A}_{13},\mathcal{A}_2\}]$$
$$-\Pr[G_k=\{\mathcal{A}_{23},\mathcal{A}_1\}]. \tag{8}$$

The sigma notation in (3) extends to amassed cliques as $\sigma_{\mathcal{A}_{i_1,i_2,\ldots,i_m}}[k]=\sigma_{\mathcal{A}_{i_1}\cup\mathcal{A}_{i_2}\cup\ldots\cup\mathcal{A}_{i_m}}[k]$ and yields:

$$\Pr[G_k=g]=\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]\sigma_{\mathcal{A}_3}[k]-\sigma_{\mathcal{M}}[k]$$
$$-(\sigma_{\mathcal{A}_{12}}[k]\sigma_{\mathcal{A}_3}[k]-\sigma_{\mathcal{M}}[k])$$
$$-(\sigma_{\mathcal{A}_{13}}[k]\sigma_{\mathcal{A}_2}[k]-\sigma_{\mathcal{M}}[k])$$
$$-(\sigma_{\mathcal{A}_{23}}[k]\sigma_{\mathcal{A}_1}[k]-\sigma_{\mathcal{M}}[k])$$
$$=\sigma_{\mathcal{A}_1}[k]\sigma_{\mathcal{A}_2}[k]\sigma_{\mathcal{A}_3}[k]-\sigma_{\mathcal{A}_{12}}[k]\sigma_{\mathcal{A}_3}[k]$$
$$-\sigma_{\mathcal{A}_{13}}[k]\sigma_{\mathcal{A}_2}[k]-\sigma_{\mathcal{A}_{23}}[k]\sigma_{\mathcal{A}_1}[k]+2\sigma_{\mathcal{M}}[k]. \tag{9}$$

The probability of a 4-clique contact graph is provided in Appendix F.

III. Contact Graph Probability Distribution

Let g be any m-clique contact graph: $g=\{\mathcal{A}_1,\mathcal{A}_2,\ldots,\mathcal{A}_m\}$. Equations (4) and (7) can be extended to compute the probability of an m-clique contact graph.

Theorem 1: The probability of an m-clique contact graph $g=\{\mathcal{A}_1,\mathcal{A}_2,\ldots,\mathcal{A}_m\}$ at discrete time k is

$$\Pr[G_k=g]=\sum_{i_1=1}^{N}\sum_{\substack{i_2=1\\ i_2\notin\{i_1\}}}^{N}\cdots\sum_{\substack{i_m=1\\ i_m\notin\{i_l\}_{l=1}^{m-1}}}^{N}\prod_{j=1}^{m}\prod_{w_u\in\mathcal{A}_j}\left(s_u[k]\right)_{i_j}. \tag{10}$$

Proof: The probability that M walkers form a contact graph $g=\{\mathcal{A}_1,\ldots,\mathcal{A}_m\}$ at discrete time k in m states $i_1,\ldots,i_m$, where $|\{i_1,\ldots,i_m\}|=m$, is equal to $\prod_{j=1}^{m}(\prod_{w_u\in\mathcal{A}_j}(s_u[k])_{i_j})$. Summing over all different m states $\{i_j\}_{j=1}^{m}$ yields the probability of the realisation g. □

Theorem 1 offers the probability of an m-clique contact graph from a combinatorial perspective. However, (10) requires to consider $\frac{N!}{(N-m)!}$ combinations of states where M walkers may form cliques $\mathcal{A}_1,\ldots,\mathcal{A}_m$, which lead to a combinatorial explosion for a large number of states N. Therefore, we derive a closed form for $\Pr[G_k=g]$.

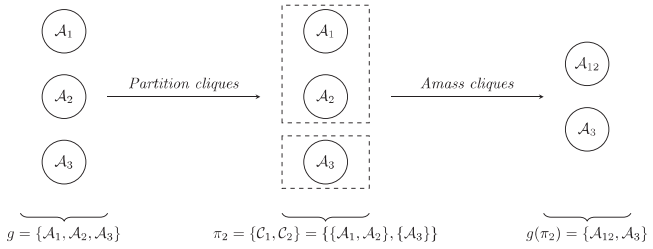

Fig. 3. Process of creating amassed clique contact graphs.

A. Amassed Clique Contact Graphs

We offer a formal definition of amassed clique graphs introduced in Section II-C3, and subsequently illustrate how the process of amassing cliques allows us to formulate our main theorem and expand (10).

Equation (8) offers insight into the recursive nature of contact graphs probabilities and partitions: *amassed clique graphs are a result of partitioning the contact graph G_k and taking the union of walkers.*

Example 1: By taking a 2-partition $\pi_2 = \{\mathcal{C}_1, \mathcal{C}_2\} = \{\{\mathcal{A}_1, \mathcal{A}_2\}, \{\mathcal{A}_3\}\}$ on the realisation $g = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$ and taking the union of cliques $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{A}_{12}$, we obtain the amassed-clique contact graph $g(\pi_2) = \{\mathcal{A}_{12}, \mathcal{A}_3\}$. Schematically, the generation of amassed clique contact graphs is shown in Fig. 3.

We call $g(\pi)$ the contact graph associated with partition π on g. Naturally, the singleton partition $\pi_3 = \{\{\mathcal{A}_1\}, \{\mathcal{A}_2\}, \{\mathcal{A}_3\}\}$ has associated contact graph $g(\pi_3) = g = \{\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3\}$. The rationale behind contact graphs generated by amassing cliques holds for any m-clique contact graph realisation $g = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\}$, thus generalising (8) to:

$$\Pr[G_k = g] = \prod_{k=1}^{m} \sigma_{\mathcal{A}_k} - \sum_{\pi \in \mathcal{P}_g^*} \Pr[G_k = g(\pi)], \tag{11}$$

where $\mathcal{P}_g$ is the set of all possible partitions on g and $\mathcal{P}_g^* = \mathcal{P}_g \backslash \{\{\mathcal{A}_1\}, \{\mathcal{A}_2\}, \ldots, \{\mathcal{A}_m\}\}$ excludes the singleton partition π_m.

We emphasize the distinction between partitions on a walker set and partitions on a contact graph. Recall from Section II-B the equivalence relationship between contact graphs and partitions on the walkers set $\mathcal{M}$. Similarly, amassed clique graphs are a special class of contact graphs, which are obtained through partitioning cliques. We denote the difference by the symbol $\mathcal{C}$ for the cells of a partition on cliques and by $\mathcal{A}$ for the cells of a partition on the walker set $\mathcal{M}$. The caveat is illustrated in Fig. 3 of Example 1, where the cells of the 2-partition π_2 are: $\mathcal{C}_1 = \{\mathcal{A}_1, \mathcal{A}_2\}, \mathcal{C}_2 = \{\mathcal{A}_3\}$ and will be used in the proof of our main theorem.

B. Main Theorem

In (11), each m-clique graph realisation $g = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ probability depends on its *associated sigma product* $\prod_{l=1}^{m} \sigma_{\mathcal{A}_l}[k]$, which allows us to reduce (11) to a closed form that depends only on sigma terms. Additionally, $\Pr[G_k = g]$ also depends on the probability of graphs associated with all partitions on g. Thus, we are motivated to reduce (11) to a closed form:

$$\Pr[G_k = g] = \sum_{\pi \in \mathcal{P}_g} \beta_m(\pi) \prod_{\mathcal{A} \in g(\pi)} \sigma_{\mathcal{A}}[k], \tag{12}$$

where $\mathcal{A}$ is a clique in the amassed clique graph $g(\pi)$ (associated with partition π on g), $\beta_m(\pi) \in \mathbb{Z}$ is the number of sigma product terms associated with contact graph $g(\pi)$ and subscript m is the number of cliques in g. We call (12) the *sigma expansion* of (11) for contact graph g. We now state our main theorem:

Theorem 2: The probability of an m-clique contact graph $g = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\}$ at discrete time k is

$$\Pr[G_k = g] = \sum_{\pi \in \mathcal{P}_g} \left(\prod_{\mathcal{C} \in \pi} (-1)^{|\mathcal{C}|-1} (|\mathcal{C}|-1)! \right) \prod_{\mathcal{A} \in g(\pi)} \sigma_{\mathcal{A}}[k], \tag{13}$$

where $|\mathcal{C}|$ denotes the number of cliques $\mathcal{A}$ in cell $\mathcal{C}$ of partition π on $g = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m\}$.

Our proof of Theorem 2 stems directly from Lemmas 1 and 2, presented below.

Lemma 1: Let $\pi_1 = \{\mathcal{M}\}$ be the 1-partition on the walker set $\mathcal{M}$. The number $\beta_m(\pi_1)$ of sigma product terms $\sigma_{\mathcal{M}}[k]$ in the sigma expansion formula (12) for the probability of an m-clique contact graph depends only on the number of cliques m as

$$\beta_m(\pi_1) = (-1)^{(m-1)}(m-1)! \tag{14}$$

Lemma 2: Let g be a m-clique contact graph. Let $\pi_q = \{\mathcal{C}_1, \ldots, \mathcal{C}_q\}$ be a q-partition on g, with $q < m$. Let the cardinality of each cell $\mathcal{C}_i$ be c_i. Let the number of sigma product terms $\prod_{i=1}^{q} \sigma_{\mathcal{C}_i}[k]$ in the sigma expansion formula of g be $\beta_m(\pi_q)$. Then

$$\beta_m(\pi_q) = \prod_{i=1}^{q} (-1)^{c_i - 1}(c_i - 1)! \tag{15}$$

Lemma 1 offers a formula for the weight $\beta_m(\pi_1)$ of the sigma product $\sigma_{\mathcal{M}}[k]$ (associated with the trivial partition $\pi_1 = \{\mathcal{M}\}$, i.e. the complete graph K_M). We build Lemma 2 from Lemma 1 as a generalisation from the trivial partition to any q-partition π_q on the walker set $\mathcal{M}$. More precisely, we find the weight $\beta_m(\pi_q)$ of the sigma product associated with any q-partition on g, where $q < m$. The proofs of Lemmas 1 and 2 are provided in Appendix E.

The proof of Theorem 2 is immediate by applying Lemma 2 to all partitions on g.

Theorem 1 considers all possible combinations of walker states in order to calculate the probability distribution of a contact graph G_k which involves performing redundant, repeated operations. The adoption of a clique-centric view and the computation of the probability of cliques via Hadamard products leads to a more efficient calculation of the probability of a contact graph realisation. Therefore, Theorem 2 offers a considerable advantage in terms of the runtime. Furthermore, because different realisations in the contact graph sample space of G_k may use the same clique probability (i.e. sigma product), we can precompute all sigma products to further speed up the calculation of the probability distribution of G_k. Both Theorem 1 and 2 thus attest the elegant mathematical tractability of the

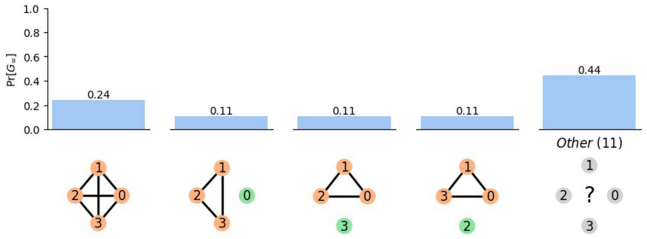

Fig. 4. Most probable 4 realisations of the contact graph G_∞ formed by 4 walkers.

RWIG model. To compare time complexities, we record the execution time of calculating the probability distribution of RWIG graphs using both (10) and (13) and present the results in Appendix D. The pseudocode for (13) in Theorem 2 is provided in Appendix C (Algorithm *RWIG-pmf*). [1]

IV. Steady-State Contact Graphs

We assume that the same $N \times N$ Markov transition matrix P, which is common for all walkers, possesses a steady-state distribution $\tilde{s}$, obeying $\tilde{s} = \tilde{s}P$. Then, the steady-state probability vector of each walker $w \in \mathcal{M}$ reduces to

$$\lim_{k\to\infty} s_w[k] = \tilde{s}. \tag{16}$$

For a clique $\mathcal{A}$ of size $|\mathcal{A}| = q$ and recalling the k-step Markov transition probability $s_j[k] = s_j[0]P^k$, taking the limit in (3) as $k \to \infty$ and invoking the existence of a steady-state in (16) yields

$$\lim_{k\to\infty} \sigma_{\mathcal{A}}[k] = \lim_{k\to\infty} \bigodot_{w_j\in\mathcal{A}} \left(s_j[0]P_j^k\right) u^T$$

$$= \left(\bigodot_{w_j\in\mathcal{A}} \tilde{s}\right) u^T = \sum_{i=1}^{N} (\tilde{s}_i)^q. \tag{17}$$

The combinatorial nature of Theorem 2 does not permit an analytical simplification of (13) in the steady state. However, (17) illustrates that cliques of the same size have the same probability, because all walkers have the same steady-state distribution $\tilde{s}$. Therefore, the probability of a steady-state contact graph does not depend on the labelling of walkers inside cliques, but rather only on clique sizes and the steady-state vector $\tilde{s}$.

Example 2: Let $M = 4$ walkers be in the steady-state $\tilde{s} = [0.1\,0.1\,0.1\,0.7]^T$. Using Theorem 2, we calculate the probability distribution of the steady-state contact graph G_∞ formed by the walkers. In Fig. 4 , we plot the most probable 4 realisations and illustrate that the second, third and fourth most probable realisations have equal probabilities and the same topologies.

Consequently, all m-clique steady-state contact graphs, which have the same m clique sizes, have equal probability and we are thus motivated to study the probability of *unlabelled* steady-state contact graphs.

[1]The code is available at: https://github.com/DavidAlmasan/rwig

A. Unlabelled Contact Graphs

Consider a steady-state m-clique contact graph realisation $g_\infty = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ with clique sizes $|\mathcal{A}_1| = q_1$, ..., $|\mathcal{A}_m| = q_m$. Additionally, denote by $\mathcal{Q} = \{q_1, \ldots, q_m\}$ a set of m positive integers which sum to M, i.e. $\sum_{i=1}^{m} q_i = M$.

We seek the number $\gamma(\mathcal{Q})$ of different steady-state m-clique contact graph realisations with M walkers where the set formed by clique sizes of each realisation is equal to the set $\mathcal{Q}$. The number γ is equal to the solution to the combinatorial problem of counting *how many ways there are to arrange M identical objects into m bins with sizes* $\{q_1, \ldots, q_m\}$. If we denote by c_j the number of cliques of size j i.e. $c_j = |\{i \in \{1, \ldots, m\}{:}q_i = j\}|$, for all $1 \le j \le M$, the solution to the problem is, by [38, equation (13.3)],

$$\gamma(\mathcal{Q}) = \frac{M!}{\prod_{i=1}^{m} q_i! \prod_{j=1}^{M} c_j!}. \tag{18}$$

Any realisation g with equal clique size set has the same structure. Hence, removing the node labels of any contact graphs with clique size set $\mathcal{Q}$ results in the *unlabelled* graph g_u which is equivalent to the set $\mathcal{Q}$. We consider unlabelled contact graphs when all walkers are found in the same steady-state because, as discussed in the beginning of Section IV, all realisations with equal clique size set have equal probability.

Unlabelled contact graphs allow us to scale RWIG to higher walker counts M by reducing the contact graph state space. For instance, Table I shows that the total number of contact graphs for $M = 9$ walkers and $N = 10$ states is 21,147. However, the total number of *unlabelled* contact graphs for $M = 9$ and $N = 10$ is only 30, which is the number of partitions of the positive integer $M = 9$ into a multiset of positive integers, such that the elements sum to M. Hence, as we increase the number of walkers M considerably above the number of states N in the Markov graph, we avoid the combinatorial Bell numbers explosion of the contact graph state space by omitting fine grained information on the walkers' clique assignment (i.e. which walker belongs to which clique) and allow for practical analysis of the clique sizes distribution.

Consider any labelled contact graph realisation g which results in an unlabelled graph g^u. Then the probability of an unlabelled graph g^u is defined by Lemma 3.

Lemma 3: The probability of a steady-state unlabelled m-clique contact graph g^u with clique sizes $\mathcal{Q} = \{q_1, \ldots, q_m\}$ and M walkers is

$$\Pr[G_\infty^u = g^u] = \gamma(\mathcal{Q}) \Pr[G_\infty = g], \tag{19}$$

where $\gamma(\mathcal{Q})$ is defined by (18), $g = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$ is a contact graph realisation obtained by any labelling of the nodes in the unlabelled realisation g^u with distinct labels from the walker set $\mathcal{M}$ and

$$\Pr[G_\infty = g] = \sum_{\pi\in\mathcal{P}_g} \left(\prod_{\mathcal{C}\in\pi} (-1)^{|\mathcal{C}|-1}(|\mathcal{C}|-1)!\right) \prod_{\mathcal{A}\in g(\pi)} \left(\sum_{i=1}^{N} \tilde{s}_i^{|\mathcal{A}|}\right) \tag{20}$$

with all walkers traverse the same Markov graph with N states and steady-state vector $\tilde{s}$.

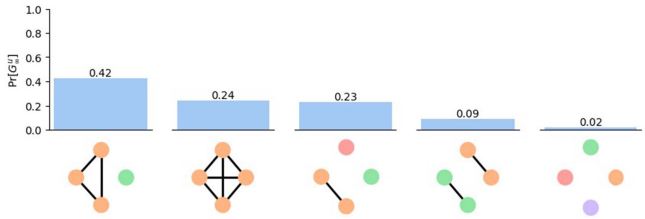

Fig. 5. Probability density of the unlabelled contact graph G^u_∞ formed by 4 walkers.

Example 3: Let $M = 4$ walkers be in the steady-state $\tilde{s} = [0.1\ 0.1\ 0.1\ 0.7]^T$. Using Lemma 3, we calculate the probability distribution of the steady-state unlabelled contact graph G^u_∞ formed by the walkers and plot it in Fig. 5 .

B. A Combinatorial Computation of the Steady-State Graph

Another way to compute the probability of a steady-state unlabelled m-clique contact graph, where each walker has the same[2] steady-state vector $\tilde{s}$, can be obtained from Theorem 1. For a labelled contact graph $g = \{\mathcal{A}_1, \ldots, \mathcal{A}_m\}$, the probability of realisation g becomes

$$\Pr[G_\infty = g] = \lim_{k\to\infty} \Pr[G_k = g] = \\ = \sum_{i_1=1}^{N} \sum_{\substack{i_2=1 \\ i_2\notin\{i_1\}}}^{N} \cdots \sum_{\substack{i_m=1 \\ i_m\notin\{i_1,\ldots,i_{m-1}\}}}^{N} \prod_{j=1}^{m} \tilde{s}_{i_j}^{q_j}. \quad (21)$$

where the clique sizes $q_i = |\mathcal{A}_i|$, for all $1 \le j \le m$, form the clique size set $\mathcal{Q} = \{q_1, \ldots, q_m\}$. The number of labelled graphs with clique size set $\mathcal{Q}$ is given by (18), and thus the probability of an unlabelled steady-state contact graph g^u with clique size set $\mathcal{Q}$ is

$$\Pr[G^u_\infty = g^u] = \frac{M! \sum_{i_1=1}^{N} \sum_{\substack{i_2=1 \\ i_2\notin\{i_1\}}}^{N} \cdots \sum_{\substack{i_m=1 \\ i_m\notin\{i_l\}_{l=1}^{m-1}}}^{N} \prod_{j=1}^{m} \tilde{s}_{i_j}^{q_j}}{\prod_{i=1}^{m} q_i! \prod_{j=1}^{M} c_j!}. \quad (22)$$

where c_j is the number of cliques of size j, for all $1 \le j \le M$.

V. Empirical Analysis

The assumption of RWIG, that all walkers found in the same Markov state at discrete-time k are connected in the contact graph, implies that the contact graph G_k is formed by the union of disconnected cliques. In this section, we analyse various empirical temporal networks to validate our assumption and we demonstrate that RWIG is able to reproduce contact graphs with similar topological properties.

A. Datasets

We inspect a series of empirical datasets collected through the SocioPatterns sensing platform (http://www.sociopatterns.org). Génois and Barrat [39] study how co-location graphs can be used as a proxy for face-to-face contacts. Similar to our fundamental assumption in RWIG, individuals are considered connected in a co-location graph if they are found to be in the same spatial location. Consequently, the co-location datasets are snapshots at discrete time steps of graphs. Our analysis of the datasets released in [39] found that all co-location samples consist of unions of disconnected cliques, in complete accord with the topology of the contact graphs generated by RWIG.

[2]If not all walker's probabilities are the same, then we must again compute all possible partitions as in Theorem 2.

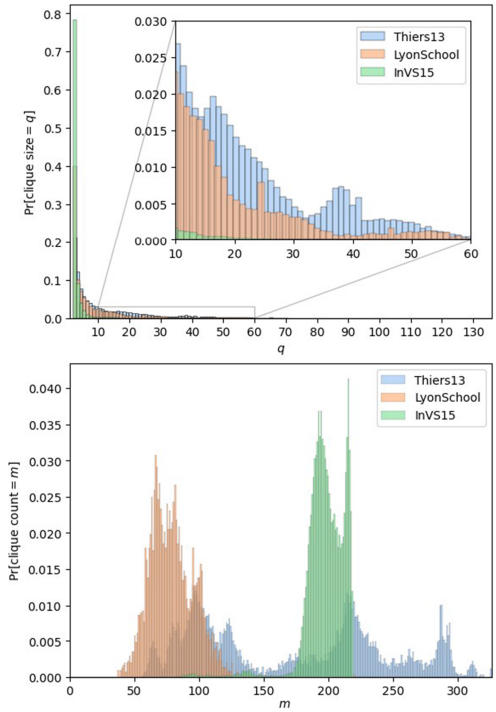

Fig. 6. Clique size distribution (top) and clique count distribution (bottom) for three co-location datasets.

We study the *clique size* distribution and the *clique count* distribution in co-location graphs. The clique size distribution is defined as the probability of observing a clique of a certain size and offers insight into possible patterns of *typical* clique sizes. The clique count quantifies the connectivity of the contact graph. Fig. 6 depicts the clique size and count distributions for three co-location datasets from [39]: InVS15 (219 nodes), LyonSchool (242 nodes) and Thiers13 (328 nodes). For the clique size distribution, we only consider cliques formed of at least two individuals.

As shown in Fig. 6, most cliques for all datasets are small in size and consist mainly of two nodes. However, the clique count distribution indicates that some datasets exhibit stronger connectivity. For instance, the co-location graphs in LyonSchool have on average fewer cliques than the graphs in InVS15 while having a larger number of nodes. Hence, there is a comparably higher propensity for larger clique sizes in the graphs from LyonSchool, which is supported by Fig. 6. Overall, there is significant variability in the structure of co-location graphs.

Finally, we investigate the impact of the number of Markov states N and the walkers' Markov transition probability matrices P_r on the accuracy of modelling empirical contact

TABLE II
STEADY-STATE VECTORS

	$\tilde{s}$
$s_N = 0.33$	[0.047 ... 0.047 0.33]
$s_N = 0.96$	[0.003 ... 0.003 0.96]
$Multimodal$	$[\frac{1}{1200}$ ... $\frac{1}{1200}$ 0.32 0.32 0.32]

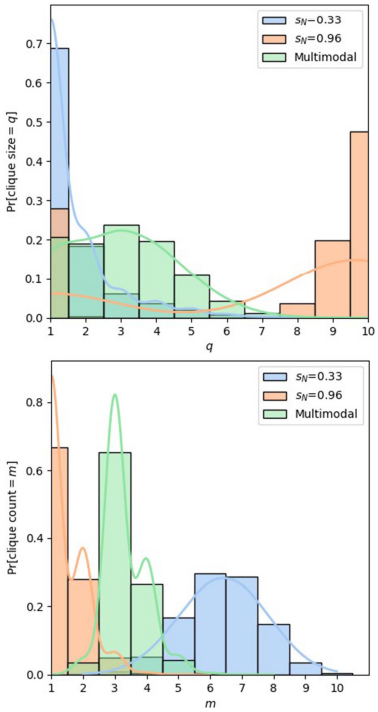

Fig. 7. Clique size distribution (top) and clique count distribution (bottom) for unlabelled steady-state contact graphs.

graph distributions using RWIG. An ablation study is found in Appendix G.

B. Simulations of Steady-State Contact Graphs

We now show how RWIG is able to generate both sparse and dense contact graphs with minimal parameter tuning. As the unlabelled steady-state contact graph distribution $\Pr[G^u]$ depends only on the steady-state distribution $\tilde{s}$, we compute the clique size and clique count distributions for a range of different steady-state distributions. We consider $M = 10$ walkers and a Markov graph with $N = 15$ states which admits a steady-state vector $\tilde{s}$. We consider three different steady-state vectors (see Table II). The first two steady-state vectors $\tilde{s} = [s_1\ s_2\ \dots\ s_N]^T$ have equal probability $s_1 = s_2 = \dots = s_{N-1}$ for the the first $N-1$ states while the probability s_N of state N takes values $s_N > s_1$ and $s_N \gg s_1$. We also consider the steady-state vector with the last three elements equal to each other $\tilde{s} = [\frac{1}{1200}\ \dots\ \frac{1}{1200}\ 0.32\ 0.32\ 0.32]$, which we call the *Multimodal* steady-state vector.

Fig. 7 illustrates the clique size and clique count distributions for $M = 10$ walkers on a $N = 15$ state Markov graph. We overlay a smooth Kernel Density Estimate (KDE) line plot on top of the histograms for better visualisation.

VI. CONCLUSION

We present RWIG, a temporal contact graph model generated by independent random walkers on a Markov graph. A random walk on a Markov graph is a realisation of a Markov process, which is specified in discrete time by a transition probability matrix P_r and an initial condition $s_r[0]$ for each walker w_r. By choosing the matrices P_r as well the vectors $s_r[0]$, any collection of discrete-time Markov processes can generate a corresponding temporal contact graph sequence consisting of disjoint cliques, which makes RWIG general.

We derive the probability distribution (Theorem 2) of the RWIG contact graphs under the assumption of known initial walker states and transition probabilities in the Markov graph. In Section V, we show that many real temporal network datasets consist of disconnected cliques, in complete accord with the graph configurations implied by the co-location principle of RWIG. We further demonstrate that RWIG is capable of producing diverse contact graphs including graphs with many small cliques (e.g. $s_N = 0.33$) or few large cliques (e.g. $s_N = 0.96$). Moreover, we illustrate how clique size variety is already imposed by only tuning the steady-state vector $\tilde{s}$.

In general, the accuracy of RWIG depends on the number of states N in the Markov graph. Hence, we perform an ablation study in Appendix G to illustrate the trade-off between model complexity and accuracy. When all walkers follow the same Markov transition probability matrix P, RWIG's accuracy in generating realistic contact graph distributions increases with the number of Markov states N. However, the additional complexity of different Markov policies is enough to instill RWIG with superior modelling performance. RWIG perfectly reproduces (i.e. with zero average error) the empirical contact graph distributions from the LyonSchool dataset even when the Markov graphs have only a few states (Appendix G, Fig. 9).

The analytical tractability of the model, along with the capability to create a wide variety of contact graphs, renders RWIG a promising basis for temporal graphs generative modelling.

VII. FURTHER WORK

We will explore extensions to the RWIG model:

First, we plan to address the inverse problem, that consists of finding the class of transition probability matrices P that generates a given K-length sequence of contact graphs $G_1, \dots, G_K$. While statistical methods such as maximum likelihood estimation lie at the heart of the problem, the complexity of the parameter search space and scarcity of similarity measures for temporal graphs make this task non-trivial.

Second, given that a link in G_k occurs, what is the probability that that link still exists at time $l > k$ in G_l? Alternatively, can RWIG's transition probability matrix be tuned to generate a "link burst" (i.e. the existence of a link over multiple time slots). Many other questions or assumptions made in the temporal graph community may be addressed from the "process point of view" of RWIG.

Third, motivated by the importance of higher-order correlations in time-respecting paths [20], [24], can RWIG be used to analytically calculate the probability of time-respecting paths of

length k? The answer would not only unravel which mechanisms (in terms of the underlying Markov graph and the transition probability matrix P) can lead to temporal graphs, whose *causal topology* –i.e. which nodes can indirectly influence each other via time-respecting paths– differs from that of the corresponding static graph, but it would also shed light on the question why many human contact patterns exhibit second-order correlations, which has been shown to strongly influence the dynamics of diffusion and epidemic spreading [13], [14].

Finally, we outline that RWIG fundamentally represents a framework for analysing the co-occurence in time and sample space of any set of possibly different stochastic processes. It would be fruitful to examine the applicability of RWIG to temporal graphs which are not inspired by human mobility. For example, if we consider the *action* space of reinforcement learning agents, RWIG could model the co-occurence in space and time of the same action taken by agents which follow possibly different policies. RWIG would thus be able to offer insight into how different agent policies converge towards optimal actions.

REFERENCES

[1] K. Y. Yap and Q.-L. Xie, "Personalizing symptom monitoring and contact tracing efforts through a COVID-19 Web-App," *Infect. Dis. Poverty*, vol. 9, 2020, Art. no. 93.

[2] G. Cencetti et al., "Digital proximity tracing on empirical contact networks for pandemic control," *Nat. Commun.*, vol. 12, 2021, Art. no. 1655.

[3] World Health Organization, "Digital tools for COVID-19 contact tracing," Geneva, Switzerland, 2020. [Online]. Available: https://who.int/publications/i/item/WHO-2019-nCoV-Contact_Tracing-Tools_Annex-2020.1

[4] J. L. Iribarren and E. Moro, "Impact of human activity patterns on the dynamics of information diffusion," *Phys. Rev. Lett.*, vol. 103, 2009, Art. no. 038702.

[5] M. Karsai et al., "Small but slow world: How network topology and burstiness slow down spreading," *Phys. Rev. EStat., Nonlinear, Soft Matter Phys.*, vol. 83, 2013, Art. no. 025102.

[6] P. Takaguchi, "Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics," *PLoS One*, vol. 8, no. 7, pp. 1–5, 2013.

[7] J.I. Perotti, H.-H. Jo, P. Holme, and J. Saramäki, "Temporal network sparsity and the slowing down of spreading," 2014, *arXiv:1411.5553*.

[8] N. Masuda, K. Klemm, and V. Eguíluz, "Temporal networks: Slowing down diffusion by long lasting interactions," *Phys. Rev. Lett.*, vol. 111, 2013, Art. no. 188701.

[9] M. Morris and M. Kretzschmar, "Concurrent partnerships and transmission dynamics in networks," *Soc. Netw.*, vol. 17, no. 3, pp. 299–318, 1995.

[10] H. H. K. Lentz, T. Selhorst, and I. M. Sokolov, "Unfolding accessibility provides a macroscopic approach to temporal networks," *Phys. Rev. Lett.*, vol. 110, 2013, Art. no. 118701.

[11] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Phys. Rev. E*, vol. 84, 2011, Art. no. 016105.

[12] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012.

[13] R. Pfitzner, I. Scholtes, A. Garas, C. J. Tessone, and F. Schweitzer, "Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks," *Phys. Rev. Lett.*, vol. 110, 2013, Art. no. 198701.

[14] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, "Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks," *Nat. Commun.*, vol. 5, 2014, Art. no. 5024.

[15] M. Rosvall, A.V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, "Memory in network flows and its effects on spreading dynamics and community detection," *Nat. Commun.*, vol. 5, 2014, Art. no. 4630.

[16] H. H. K. Lentz et al., "Disease spread through animal movements: A static and temporal network analysis of pig trade in Germany," *PLoS One*, vol. 11, no. 5, 2016, Art. no. e0155196.

[17] J. Peng et al., "Signal propagation in complex networks," *Phys. Rep.*, vol. 1017, pp. 1–96, 2023, doi: 10.1016/j.physrep.2023.03.005.

[18] D. Hou, C. Gao, Z. Wang, X. Li, and X. Li, "Random full-order-coverage based rapid source localization with limited observations for large-scale networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 5, pp. 4213–4226, Sep./Oct. 2024, doi: 10.1109/TNSE.2024.3406394.

[19] F. Bois and G. Gayraud, "Probabilistic generation of random networks taking into account information on motifs occurrence," *J. Comput. Biol.*, vol. 22, no. 1, pp. 25–36, 2015.

[20] I. Scholtes, "When is a network a network? Multi-order graphical model selection in pathways and temporal networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2017, vol. 17, pp. 1037–1046.

[21] D. Zhou, L. Zheng, J. Han, and J. He, "A data-driven graph generative model for temporal interaction networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 401–411.

[22] G. Zeno, L. Fond, and T. Neville, "DYMOND: Dynamic motif-nodes network generative model," in *Proc. Web Conf.*, 2021, pp. 718–729.

[23] A. Longa, G. Cencetti, S. Lehmann, A. Passerini, and B. Lepri, "Generating fine-grained surrogate temporal networks," *Commun. Phys.*, vol. 7, 2024, Art. no. 22.

[24] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," *Nat. Phys.*, vol. 15, pp. 313–320, 2019.

[25] S. Shvydun and V. P. Mieghem, "System identification for temporal networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 2, pp. 1885–1895, Mar./Apr. 2024.

[26] H. Barbosa et al., "Human mobility: Models and applications," *Phys. Rep.*, vol. 734, pp. 1–74, 2018.

[27] B. Chang et al., "Markov modulated process to model human mobility," *Stud. Comput. Intell.*, vol. 1072, pp. 607–618, 2022.

[28] A. Panisson, A. Barrat, C. Cattuto, V. W. den Broeck, W. G. Ruffo, and R. Schifanella, "On the dynamics of human proximity for data diffusion in ad-hoc networks," *Ad Hoc Netw.*, vol. 10, no. 8, pp. 1532–1543, 2012.

[29] G. Mauro, M. Luca, A. Longa, B. Lepri, and L. Pappalardo, "Generating mobility networks with generative adversarial networks," *EPJ Data Sci.*, vol. 11, 2022, Art. no. 58.

[30] X. Kui, A. Samanta, X. Zhu, S. Zhang, Y. Li, and P. Hui, "Energy-aware temporal reachability graphs for time-varying mobile opportunistic networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9831–9844, Oct. 2018.

[31] N. Huynh and J. Barthelemy, "A comparative study of topological analysis and temporal network analysis of a public transport system," *Int. J. Transp. Sci. Technol.*, vol. 11, no. 2, pp. 392–405, 2022.

[32] A. P. Riascos and D. P. Sanders, "Mean encounter times for multiple random walkers on networks," *Phys. Rev. E*, vol. 103, 2021, Art. no. 042312.

[33] M. Bestehorn, A. Riascos, T. Michelitsch, and B. Collet, "A Markovian random walk model of epidemic spreading," *Continuum Mech. Thermodyn.*, vol. 33, pp. 1207–1221, 2021.

[34] N. Masuda, M. Porter, and R. Lambiotte, "Random walks and diffusion on networks," *Phys. Rep.*, vol. 716, pp. 1–58, 2017.

[35] V. P. Mieghem, *Graph Spectra for Complex Networks*. 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 2023.

[36] V. P. Mieghem, *Performance Analysis of Complex Networks and Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[37] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*. New York, NY, USA: Dover Publications, 1968.

[38] V. Lint, J. H., and R. M. Wilson, *A Course in Combinatorics*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[39] M. Génois and A. Barrat, "Can co-location be used as a proxy for face-to-face contacts?," *EPJ Data Sci.*, vol. 7, no. 11, pp. 1–18, 2018.

Anton-David Almasan received the Master of Engineering degree from the University of Cambridge, Cambridge, U.K., with specialisation in information and computer engineering. Since 2024, he has been working toward the Ph.D. degree with the Delft University of Technology, Delft, The Netherlands. Before joining Delft, he was an AI Consultant with Deepsea Technologies and a Data Scientist with Thales U.K. His research interests mainly include machine learning, network science, and stochastic processes.

Sergey Shvydun received the master's degree (with distinction) in business informatics and the Ph.D. degree in applied mathematics *(cum laude)* from HSE University, Moscow, Russia, in 2014 and 2020, respectively. He was an Associate Professor with the HSE University and as a Senior Research Fellow with the Institute of Control Sciences, Russian Academy of Science, Moscow. Since 2023, he has been a Postdoctoral Researcher with the Delft University of Technology, Delft, The Netherlands. He is the author of more than 30 papers in peer-reviewed journals and edited volumes and one book on centrality in networks. His research interests mainly include network science, machine learning, operations research, and social choice theory.

Ingo Scholtes received the Diploma (with Distinction) and the Doctorate degree in computer science *(summa cum laude)* from the University of Trier, Trier, Germany, in 2005 and 2011, respectively. From 2011 to 2018, he was a Postdoctoral Researcher and Lecturer with the Chair of Systems Design, ETH Zürich, Zürich, Switzerland. From 2019 to 2021, he was a Professor of data analytics with the University of Wuppertal, Wuppertal, Germany, and was also with the University of Zurich from 2018 to 2024. In 2021, he was a Chaired Professor with the University of Würzburg. He is currently a Professor of machine learning for complex networks with the Center for Artificial Intelligence and Data Science (CAIDAS), University of Würzburg, Würzburg, Germany, and SNF Professor of data analytics with the Department of Informatics (IfI), University of Zurich, Zurich, Switzerland. He is the author of more than 80 papers in peer-reviewed conference proceedings and interdisciplinary journals. Apart from methodological contributions with the interface between network science and machine learning, his research interests mainly include network science, modelling and analysis of temporal networks, applications of machine learning to complex networks, applications of (temporal) graph learning in collaborative software engineering, single cell biology, and astrophysics. He is also the Vice-Spokesperson with the Center for Artificial Intelligence and Data Science (CAIDAS) and Founding Co-Chair with the topical section Computational Social Science, German Informatics Society (GI e.V.). He was on the Editorial Board of *Advances in Complex Systems* and is an Associate Editor for *EPJ Data Science*. In 2014, he was the recipient of the Junior-Fellowship from the German Informatics Society (GI e.V.) and SNSF-Professorship from the Swiss National Science Foundation in 2018.

Piet Van Mieghem (Fellow, IEEE) received the master's and Ph.D. degrees in electrical engineering from the K.U. Leuven, Leuven, Belgium, in 1987 and 1991, respectively. He is currently a Professor with the Delft University of Technology, Delft, The Netherlands, and has been the Chairman with Section Network Architectures and Services (NAS) since 1998. He is the author of four books which include *Performance Analysis of Communications Networks and Systems*, *Data Communications Networking*, *Graph Spectra for Complex Networks*, and *Performance Analysis of Complex Networks and Systems*. From 1987 to 1991, he was with Interuniversity Micro Electronic Center (IMEC). From 1993 to 1998, he was a Member with Alcatel Corporate Research Center, Antwerp, Belgium. From 1992 to 1993, he was a Visiting Scientist with the Massachusetts Institute of Technology, Cambridge, MA, USA, Visiting Professor with the University of California, Los Angeles, Los Angeles, CA, USA, in 2005, and was also with Cornell University, Ithaca, NY, USA, in 2009, Stanford University, Stanford, CA, USA, in 2015, and Princeton University, Princeton, NJ, USA, in 2022. He is also on the Editorial Board of the OUP *Journal of Complex Networks*. He was the recipient of the Advanced ERC Grant 2020 for ViSiON, Virus Spread in Networks. From 2005 to 2006, he was member of the Editorial Board of *Computer Networks*, IEEE/ACM TRANSACTIONS ON NETWORKING from 2008 to 2012, *Journal of Discrete Mathematics* from 2012 to 2014, and *Computer Communications* from 2012 to 2015. He is also on a Board Member with the Netherlands Platform of Complex Systems, a Steering Committee Member with Dutch Network Science Society, External Faculty Member with the Institute for Advanced Study (IAS), University of Amsterdam, Amsterdam, The Netherlands.