9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Randomized reference models for temporal networks

Keywords: random network models, temporal networks, data analysis, statistical inference, numerical methods

Extended Abstract

Empirical social networks and dynamic processes that take place in these situations show heterogeneous, non-Markovian, and intrinsically correlated topologies and dynamics. This makes their analysis particularly challenging. Randomized reference models (RRMs) have emerged as a general and versatile toolbox for studying such systems [1, 2]. Defined as random networks with given features constrained to match those of an input (empirical) network, they are notably used as null models for hypothesis testing and, more generally, to investigate the relationship between different network features and their roles in dynamic phenomena.

RRMs are typically implemented as procedures that reshuffle an empirical network, making them very generally applicable. However, while a multitude of different randomization techniques are found in the literature [1, 2], the effects of most shuffling procedures on network features remain poorly understood, rendering their use non-trivial and susceptible to misinterpretation. For example, the lack of unified naming conventions for RRMs has lead to a situation where the algorithms producing equivalent RRMs are given a multitude of different names, and possibly worse, where multiple algorithms producing different RRMs are given the same names [3].

We propose a unified framework for the important class of RRMs generated by uniform shuffling procedures, which by analogy to statistical physics we name microcanonical RRMs (MRRMs) [4]. An MRRM constrains a chosen vector of features, $\mathbf{x}$, to take exactly the same value as in the empirical network G^* but generates network that are otherwise maximally random, i.e., it samples networks G with probability:

$$P_{\mathbf{x}}(G|G^*) = \frac{\delta_{\mathbf{x}(G),\mathbf{x}(G^*)}}{\Omega_{\mathbf{x}}(G^*)},$$

where δ is the Kronecker delta function, and $\Omega_{\mathbf{x}}(G^*) = \sum_G \delta_{\mathbf{x}(G),\mathbf{x}(G^*)}$ is a normalization constant.

We use this framework to propose a concise naming convention that completely and unambiguously describes an MRRM: an MRRM is named as P[$\mathbf{x}$], where $\mathbf{x}$ is the list of features it constrains. For example, the Erdős-Rényi model, which constrains the number of links L, is named P[L], while the configuration model, which constrains the degree sequence $\mathbf{k}$, is named P[$\mathbf{k}$].

Our framework furthermore enables us to formally compare and order MRRMs in terms of their randomness, where one MRRM, P[$\mathbf{z}$], is more random than another, P[$\mathbf{x}$], if the ensemble of networks generated by the former always contains the ensemble generated by the latter (Fig. 1A). We denote this by P[$\mathbf{x}$] $\leq$ P[$\mathbf{z}$]. We show that MRRMs correspond to partitions of the state space of possible networks, so the ordering can be understood as P[$\mathbf{z}$] being a coarser partition than P[$\mathbf{x}$] (Fig. 1B). To compare MRRMs in practice, we show that if one feature can be written as a function of another, $\mathbf{z} = \mathbf{f}(\mathbf{x})$, then P[$\mathbf{x}$] $\leq$ P[$\mathbf{z}$]. For example, the number of links can be obtained as a function of the degree sequence, $L = \sum_i k_i/2$, so P[L] is more random than P[$\mathbf{k}$] (Fig. 1C).

We additionally show when and how we may generate new MRRMs by composition of existing ones (Fig. 1D,E), allowing to generate hundreds of new MRRMs from the existing ones surveyed in the literature.

Focusing on temporal networks, we apply our framework to build a taxonomy of MRRMs that orders them and deduces their effects on important network features (Fig. 1F). We finally apply our framework to unravel the influence of different features of an empirical network of mobile-phone calls on the spread of information in the network (Fig. 1G).

Our taxonomy provides a reference for the use of MRRMs, and the theoretical foundations laid by the framework may further serve as a base for the development of a principled and automatized way to generate and apply randomized reference models for the study of networked systems.

References

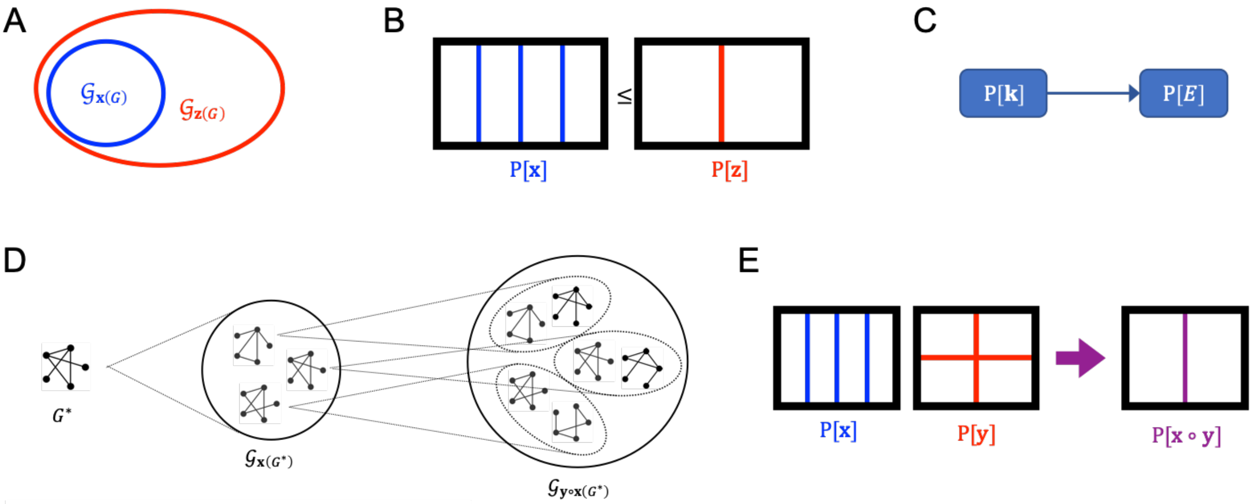
[1] Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012). 1108.1780.

[2] Holme, P. Modern temporal network theory: a colloquium. *Eur. Phys. J. B* **88**, 1–30 (2015).

[3] Holme, P. & Saramäki, J. A map of approaches to temporal networks. In *Temporal Network Theory*, 1–24 (Springer, 2019).

[4] Presse, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).

[5] Wu, Y., Zhou, C., Xiao, J., Kurths, J. & Schellnhuber, H. J. Evidence for a bimodal distribution in human communication. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18803–18808 (2010).



F

Canonical name	Common name	$\mathcal{L}$	s	w	A	Θ	$\Delta\tau$	t^1
P[0]	*Identity permutation*	✓	✓	✓	✓	✓	✓	✓
P[$\mathbf{w}$,$p(\mathbf{\Theta})$]	*Weight-constrained link shuffling*	✓	✓	✓	✓	p	p	p
P[$\mathcal{L}$,$p(\mathbf{\Theta})$]	*Topology-constrained link shuffling*	✓	μ	p	✓	p	p	p
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$,$\mathbf{t}^1$]	*Inter-event shuffling*	✓	✓	✓	μ	✗	p	✓
P[$\mathbf{w}$,$\mathbf{t}$]	*Timestamp shuffling*	✓	✓	✓	✓	✗	✗	✗
P[$\mathbf{w}$]	*Weight-constrained timeline shuffling*	✓	✓	✓	μ	✗	✗	✗
P[$\mathcal{L}, p(\mathbf{w}), \mathbf{t}$]	Composition of P[$\mathcal{L}, p(\mathbf{\Theta})$] and P[$\mathbf{w}, \mathbf{t}$]	✓	μ	p	✓	✗	✗	✗
P[$\mathcal{L}, p(\mathbf{w})$]	Composition of P[$\mathcal{L}, p(\mathbf{\Theta})$] and P[$\mathbf{w}$]	✓	μ	p	μ	✗	✗	✗

Features: $\mathcal{L}$: static topology; **s**: node strengths; **w**: link weights; **A**: cumulative activity at each time t; $\mathbf{\Theta}$: time-ordering of events between each pair of nodes; $\boldsymbol{\Delta\tau}$: inter-event intervals; $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$: distribution of inter-event intervals on each individual link; $\mathbf{t}^1$: first contact on each link.
Symbols: ✓ feature completely conserved; p distribution (i.e. the multiset) of individual values in sequence conserved; μ mean value of the individual features in sequence conserved; ✗ feature not conserved.

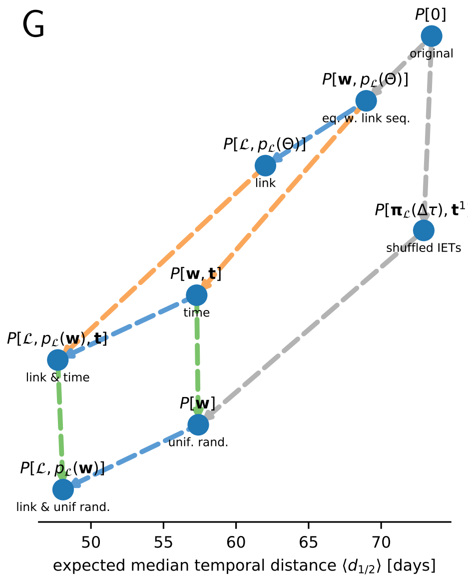


Figure 1: **A,** Two MRRMs are comparable and P[$\mathbf{x}$] $\leq$ P[$\mathbf{z}$] if the ensemble $G_{\mathbf{x}(G)}$ generated by P[$\mathbf{x}$] is always contained in the ensemble $G_{\mathbf{z}(G)}$ generated by P[$\mathbf{z}$]. **B,** In terms of partitions, P[$\mathbf{x}$] $\leq$ P[$\mathbf{z}$] means that P[$\mathbf{x}$] is finer than P[$\mathbf{z}$]. **C,** Hasse diagram depicting the hierarchy of the Erdős-Rényi model P[L] and the configuration model P[$\mathbf{k}$] (a link from one node to another means that the former is less random than the latter). **D,** Many new MRRMs can be generated from existing, compatible MRRMs by composition, consisting in applying one MRRM to the outputs of the other. **E,** The composition P[$\mathbf{x}, \mathbf{y}$] is the least random MRRM that is more random than both (their least upper bound in terms of partitions). **F,** Effects of a hierarchy of MRRMs on selected temporal network features (hierarchical relationships between MRRMs shown in G). The bottom two MRRMs are obtained by composition of two other MRRMs, i.e., by applying one MRRM to the outputs of the other. **G,** Application of the selected MRRMs to understand how different network features influence the expected median temporal distance $\langle d_{1/2}\rangle$ (average time to infect half of the nodes for the fastest possible spreading process) in a temporal mobile phone communication network (from [5]). The x-axis position of a circle shows $\langle d_{1/2}\rangle$ of the MRRM applied to the empirical network. Links between nodes means that the corresponding MRRMs can be formally compared and that the lower is more random than the upper. Absolute y-axis positions are arbitrary and not indicative of how random the MRRMs are; only MRRMs linked by arrows can be formally compared. Colored links indicate that the same features were removed: **t** for green links, $\mathbf{w} \hookrightarrow \mathscr{L}, p_{\mathscr{L}}(\mathbf{w})$ for blue links, and $p_{\mathscr{L}}(\Theta) \hookrightarrow \mathbf{t}$ for orange links. Combination effects are almost absent, permitting a simple summary of the results: the typical temporal distance in the data is around 73 days and in the most random MRRM applied here around 48 days. Out of that difference, around 12-14 days is explained by link-activation-sequence features, 7-9 days by weight-topology correlations, and 4 days by link-timeline-topology correlations.