Microcanonical randomised reference models for temporal networks

Aalto University

Mikko Kivelä

HONS@NetSci'19 2019-05-28

Typical network data analysis

Compute statistics of topology
and/or dynamics of the network

Compare to random network models

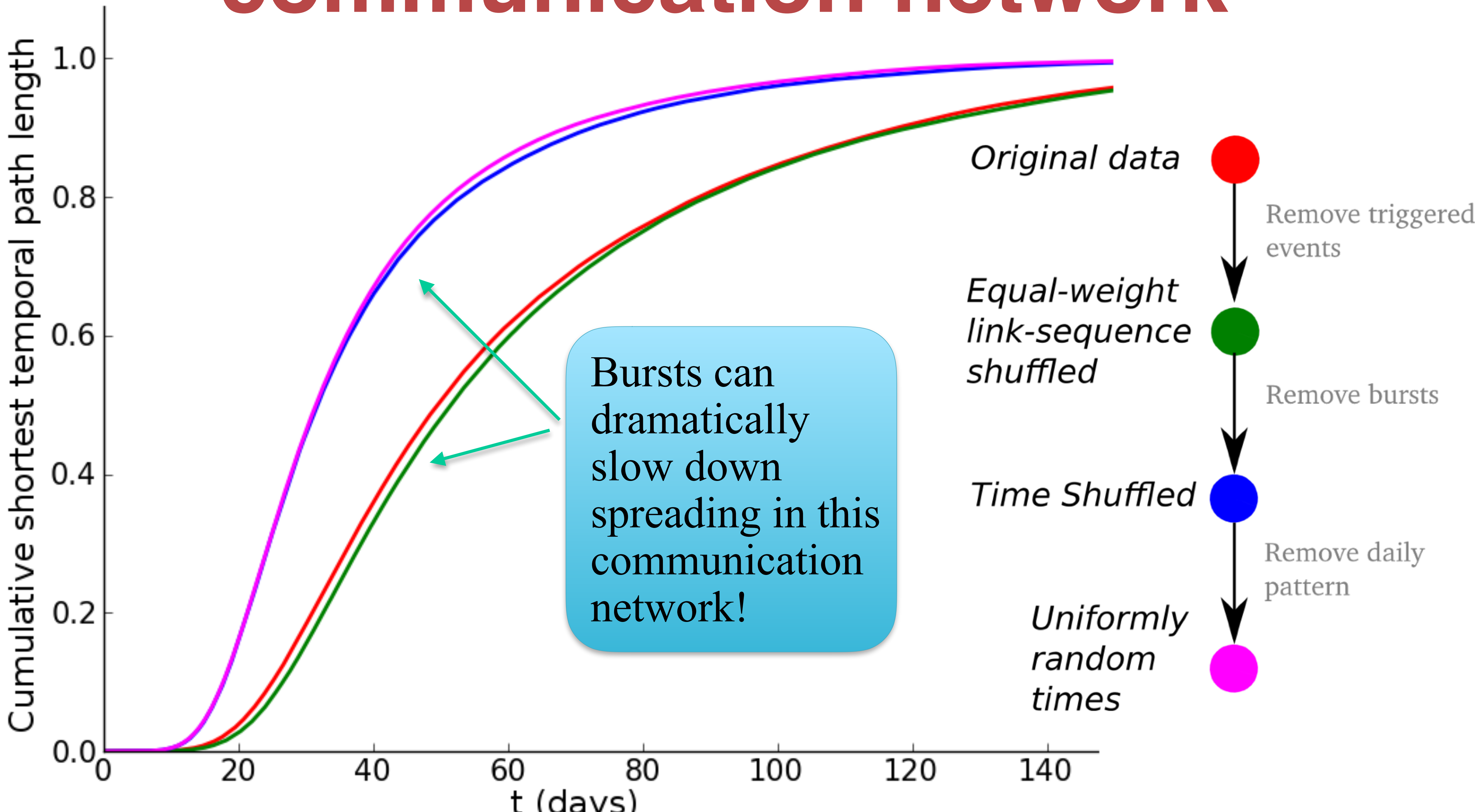
Is it so for networks with the same density?
→
Compare to *Erdős-Rényi model*

Is it so because of the degree distribution?
→
Compare to *configuration model*

What are the “Erdős-Rényi models” and “configuration models” of higher-order networks?

What are the “Erdős-Rényi models” and “configuration models” of temporal networks?

Temporal path lengths in the communication network



M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, J. Saramäki, M. Karsai: *Multiscale analysis of spreading in a large communication network*, J. Stat. Mech. 3 P03005 (2012)
M Karsai, M Kivelä, RK Pan, K Kaski, J Kertész, AL Barabási, J Saramäki: Small but slow world: How network topology and burstiness slow down spreading, Physical Review E 83 (2), 025102 (2011)

Randomized reference models for temporal networks

L. Gauvin,[1] M. Génois,[2] M. Karsai,[3] M. Kivelä,[4] T. Takaguchi,[5] E. Valdano,[6] and C. L. Vestergaard[2,7,8,*]

“randomizations”

“null models”

“strategies”

“reshuffling methods”

“methods”

“schemes”

“procedures”

“randomization techniques”

arXiv:1806.04032

Randomized reference models for temporal networks

L. Gauvin,[1] M. Génois,[2] M. Karsai,[3] M. Kivelä,[4] T. Takaguchi,[5] E. Valdano,[6] and C. L. Vestergaard[2,7,8,*]

"random time shuffle"

"time-shuffling"

"permuted times"

"shuffled time stamps"

"random dynamic"

"randomly permuted times"

"reconfigure"

arXiv:1806.04032

Randomized reference models for temporal networks

L. Gauvin,[1] M. Génois,[2] M. Karsai,[3] M. Kivelä,[4] T. Takaguchi,[5] E. Valdano,[6] and C. L. Vestergaard[2,7,8,*]

"random time sh

…e-shuffling"

"shuffled time st

…n dynamic"

"randomly p

"reconfigure"

These are all the **same** method/model!

Many more exist in the literature

arXiv:1806.04032

Applications

Systems

- Communication (email, sms, calls, …)
- Contact (physical proximity, sexual partners, …)
- Mobility (airline, public transport in cities, …)
- Transportation (cattle movements, cargo ships, …)
- …

Measurements

- Infection spreading: SI, SIR, …
- Voter model
- Complex contagion
- Motif analysis
- …

Microcanonical random reference models (MRRMs)

- Constraint/feature **x**(G) = function with network as input
- MRRM = sample uniformly randomly out of all networks with the same value of **x**(G*) as the input network G*

1, **x**(G)=**x**(G*)
0, otherwise

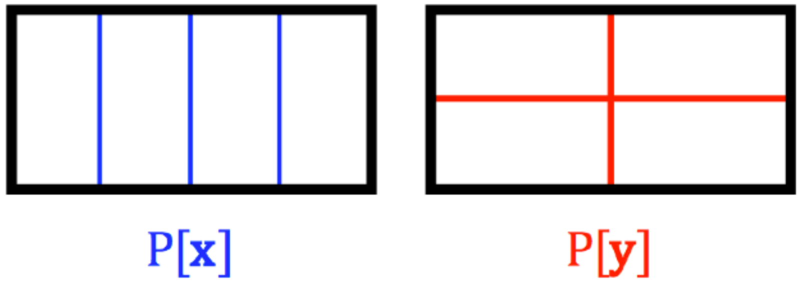
$$P_{\mathbf{x}}(G|G^*) = \frac{\delta_{\mathbf{x}(G),\mathbf{x}(G^*)}}{\Omega_{\mathbf{x}}(G^*)}$$

# of G's for which **x**(G)=**x**(G*)

- Examples, ER model: constraint is the number of edges; configuration model: constraint is the degree distribution

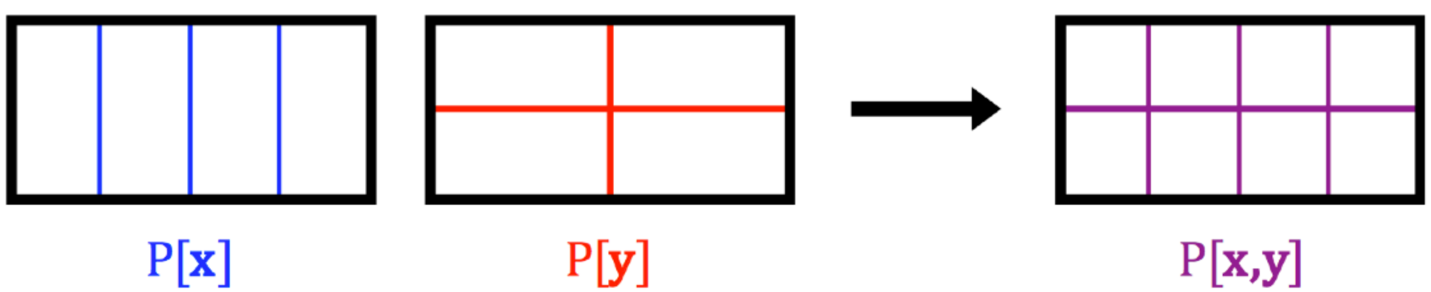
MRRMs, equivalent representations

1. A shuffling/sampling algorithm P[**x**]
2. The constraint function **x**
3. Partition of the space of all temporal networks (two networks in the same part if **x** gets the same value for them)



Microcanonical random reference models (MRRMs)

- Almost all random reference models we found in the literature are MRRMs
- Constraints can be added together to form new constraints
    - (**x**,**y**)(G)=(**x**(G),**y**(G))

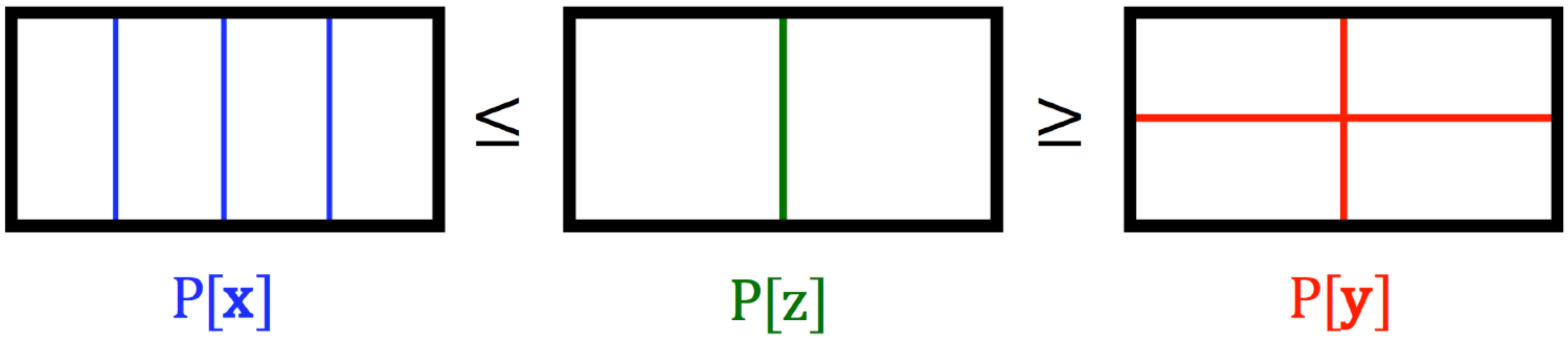


- Suggest a canonical naming convention based on the features

Shuffling method in the literature

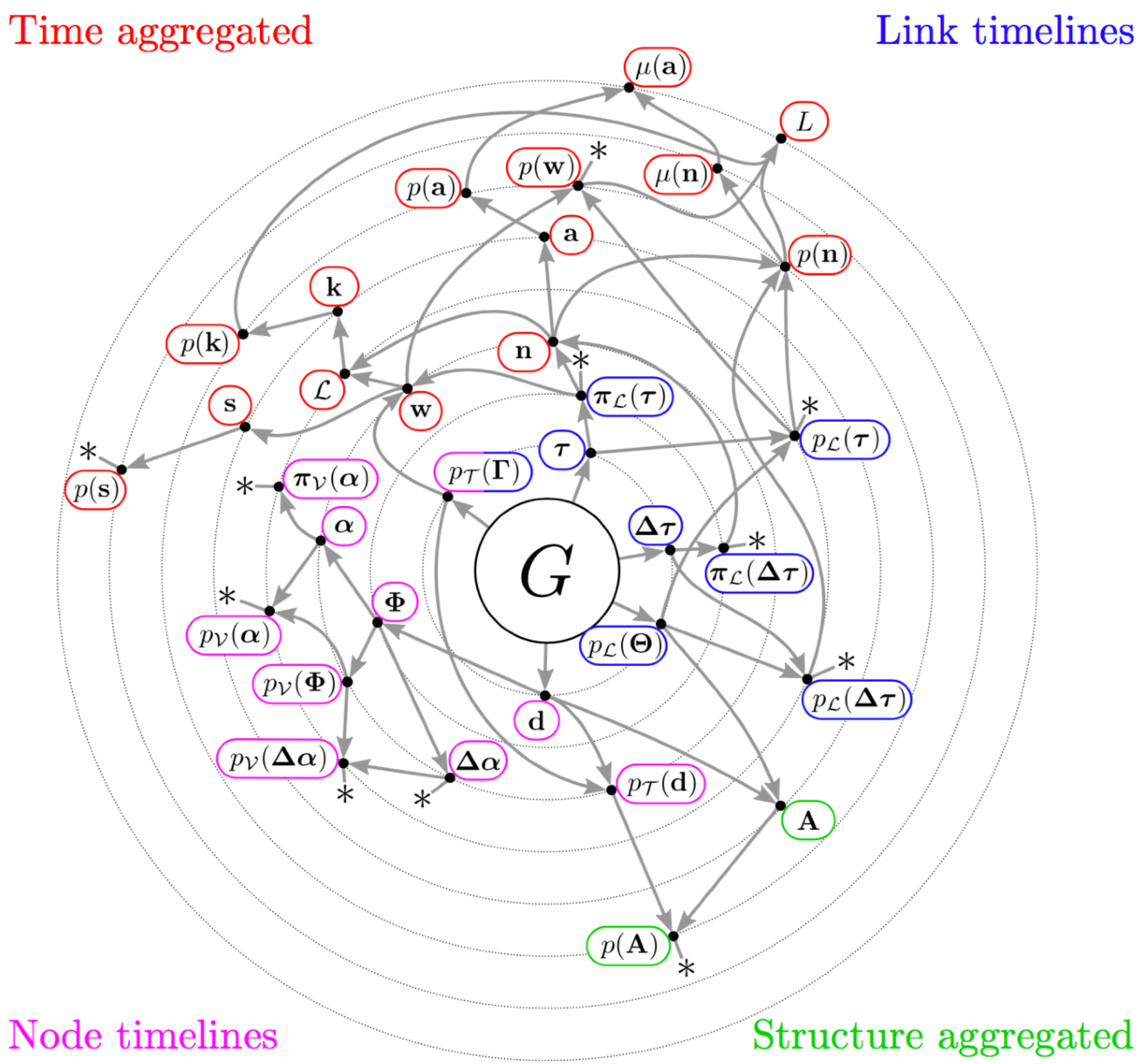
Canonical name	Common name	Aggregated								Temporal-topological									
		topological			weighted				temp.	node					link				
		G^{stat}	k_i	L	a_i	s_i	n_ℓ	w_ℓ	E^t	g^m	Φ_i	α_i^m	$\Delta\alpha_i^m$	d_i^t	Θ_ℓ	τ_ℓ^m	$\Delta\tau_\ell^m$	t_ℓ^1	t_ℓ^w
Event shufflings:																			
$\mathrm{P}[E]$	*Event shuffling*	–	–	–	–	μ	–	–	μ	–	–	–	–	μ	–	–	–	–	–
$\mathrm{P}[E,G^{\text{stat}}]$		x	x	x	–	μ	–	μ	μ	–	–	–	–	μ	–	–	–	–	–
Link shufflings:																			
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta})]$	*Link shuffling*	–	μ	x	μ	μ	p	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbb{I}_\lambda]$		$\mathbb{I}_\lambda$	μ	x	μ	μ	p	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbf{k}]$	*Maslov-Sneppen*	–	x	x	μ	μ	p	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbb{I}_\lambda,\mathbf{k}]$		$\mathbb{I}_\lambda$	x	x	μ	μ	p	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),G^{\text{stat}}]$		x	x	x	μ	μ	p	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbf{w}]$		x	x	x	μ	x	p	x	x	–	–	–	–	$\mu_{\mathcal{T}}$	$p_{(\ell)}$	$p_{(\ell)}$	$p_{(\ell)}$	p	p
Timeline shufflings:																			
$\mathrm{P}[\mathbf{w}]$	*Timeline shuffling*	x	x	x	–	x	–	x	μ	–	–	–	–	μ	–	–	–	–	–
$\mathrm{P}[\mathbf{w},\mathbf{t}^1,\mathbf{t}^w]$		x	x	x	–	x	–	x	μ	–	–	–	–	μ	–	–	$\mu_{\mathcal{L}}$	x	x
$\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{\Delta\tau})]$		x	x	x	x	x	x	x	μ	–	–	–	–	μ	–	$\pi_{\mathcal{L}}$	$\pi_{\mathcal{L}}$	–	–
$\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{\Delta\tau}),\mathbf{t}^1]$		x	x	x	x	x	x	x	μ	–	–	–	–	μ	–	$\pi_{\mathcal{L}}$	$\pi_{\mathcal{L}}$	x	x
$\mathrm{P}[\mathbf{per}(\mathbf{\Theta})]$		x	x	x	x	x	x	x	μ	–	–	–	–	μ	–	x	x	–	–
Sequence shufflings:																			
$\mathrm{P}[p_{(m)}(\mathbf{g})]$	*Sequence shuffling*	x	x	x	–	x	–	x	p	$p_{(m)}$	–	–	–	$p_{(m)}$	–	–	–	–	–
$\mathrm{P}[p_{(m)}(\mathbf{g}),\mathbf{sgn}(\mathbf{E})]$		x	x	x	–	x	–	x	p,sgn	$p_{(m)}$	–	–	–	$p_{(m)}$	–	–	–	–	–
Snapshot shufflings:																			
$\mathrm{P}[\mathbf{E}]$	*Snapshot shuffling*	–	–	–	–	μ	–	–	x	–	–	–	–	$\mu_{\mathcal{T}}$	–	–	–	–	–
$\mathrm{P}[\mathbf{E},\mathbf{\Phi}]$		–	–	–	–	μ	–	–	x	–	x	x	x	$\mu_{\mathcal{T}}$	–	–	–	–	–
$\mathrm{P}[\mathbf{d}]$		–	–	–	–	μ	–	–	x	–	x	x	x	x	–	–	–	–	–
$\mathrm{P}[\mathbf{iso}(\mathbf{g})]$		–	–	–	–	μ	–	–	x	$\simeq$	–	–	–	$\pi_{\mathcal{T}}$	–	–	–	–	–
$\mathrm{P}[\mathbf{iso}(\mathbf{g}),\mathbf{\Phi}]$		–	–	–	–	μ	–	–	x	$\simeq$	x	x	x	$\pi_{\mathcal{T}}$	–	–	–	–	–
$\mathrm{P}[\mathbf{E},G^{\text{stat}}]$		x	x	x	–	μ	–	μ	x	–	–	–	–	$\mu_{\mathcal{T}}$	–	–	–	–	–
Timeline and snapshot shufflings:																			
$\mathrm{P}[\mathbf{w},\mathbf{E}]$	*Time-stamp shuffling*	x	x	x	–	x	–	x	x	–	–	–	–	$\mu_{\mathcal{T}}$	–	–	–	–	–
Compositions:																			
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta})]\mathrm{P}[E,G^{\text{stat}}]$		–	μ	x	–	μ	–	μ	μ	–	–	–	–	μ	–	–	–	–	–
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbf{k}]\mathrm{P}[\mathbf{w},\mathbf{E}]$		–	x	x	–	μ	–	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	–	–	–	–	–
$\mathrm{P}[p_{(\ell)}(\mathbf{\Theta}),\mathbb{I}_\lambda,\mathbf{k}]\mathrm{P}[\mathbf{w},\mathbf{E}]$		$\mathbb{I}_\lambda$	x	x	–	μ	–	p	x	–	–	–	–	$\mu_{\mathcal{T}}$	–	–	–	–	–

Hierarchy/partial order



- Partial order: **x** < **z** if partition of **x** is refinement of partition of **z**
  - **x** is more strict constraint than **z**
  - P[**z**] "shuffles more" than P[**x**]
  - Not all MRRMs comparable (e.g., P[**x**] and P[**y**])
- Example: ER model "shuffles more than" the configuration model

Shuffled constraints in the literature



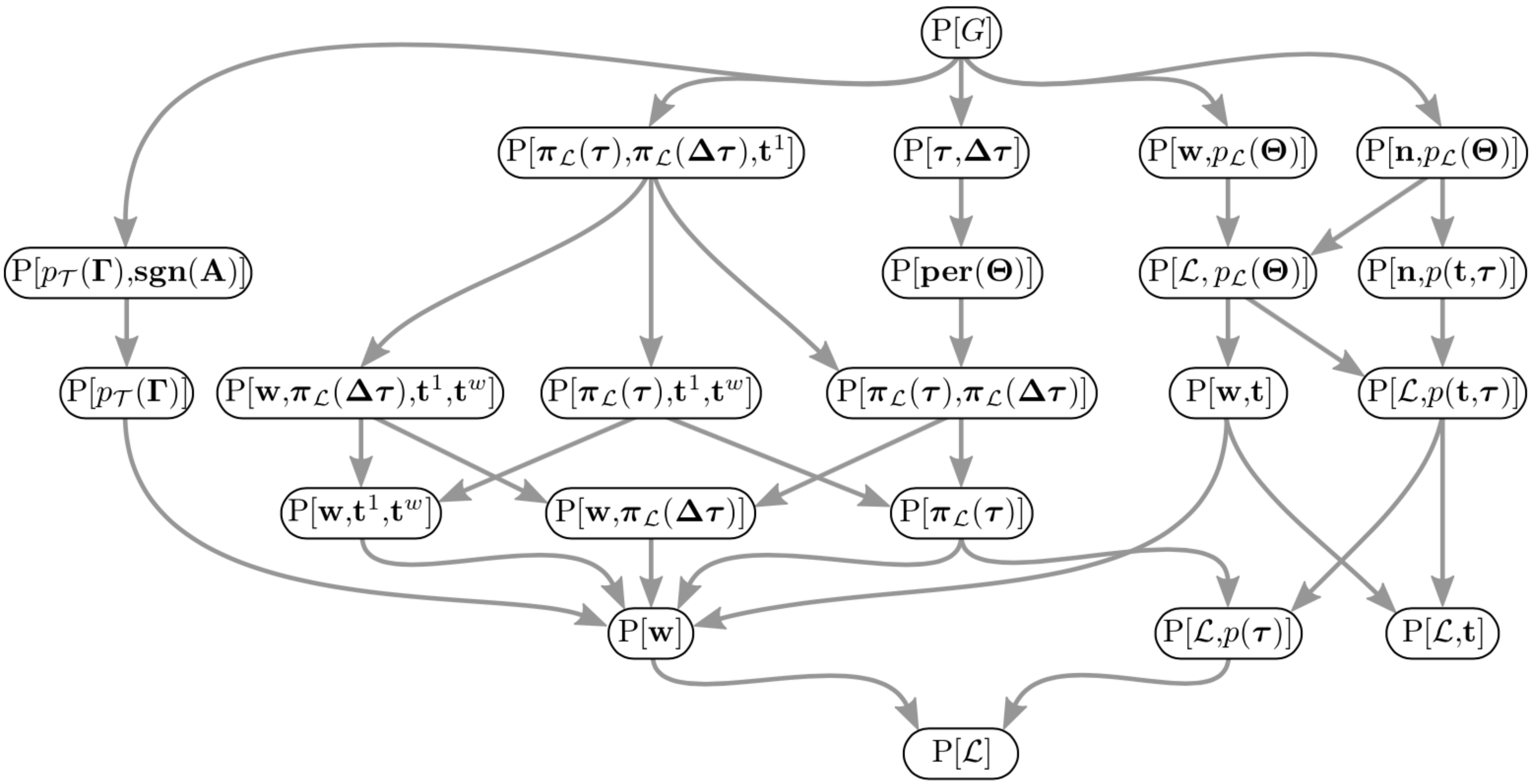
P[**x**] → P[**y**]

=

P[**x**] < P[**y**]

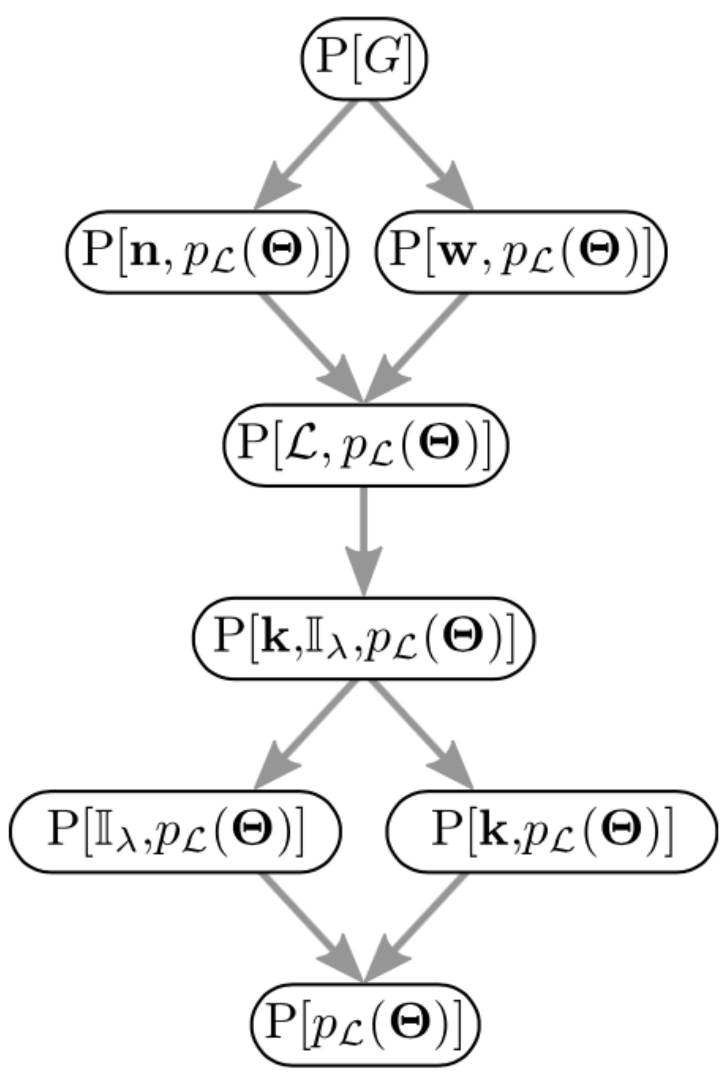
Shuffling methods in the literature

Link timeline shuffling methods

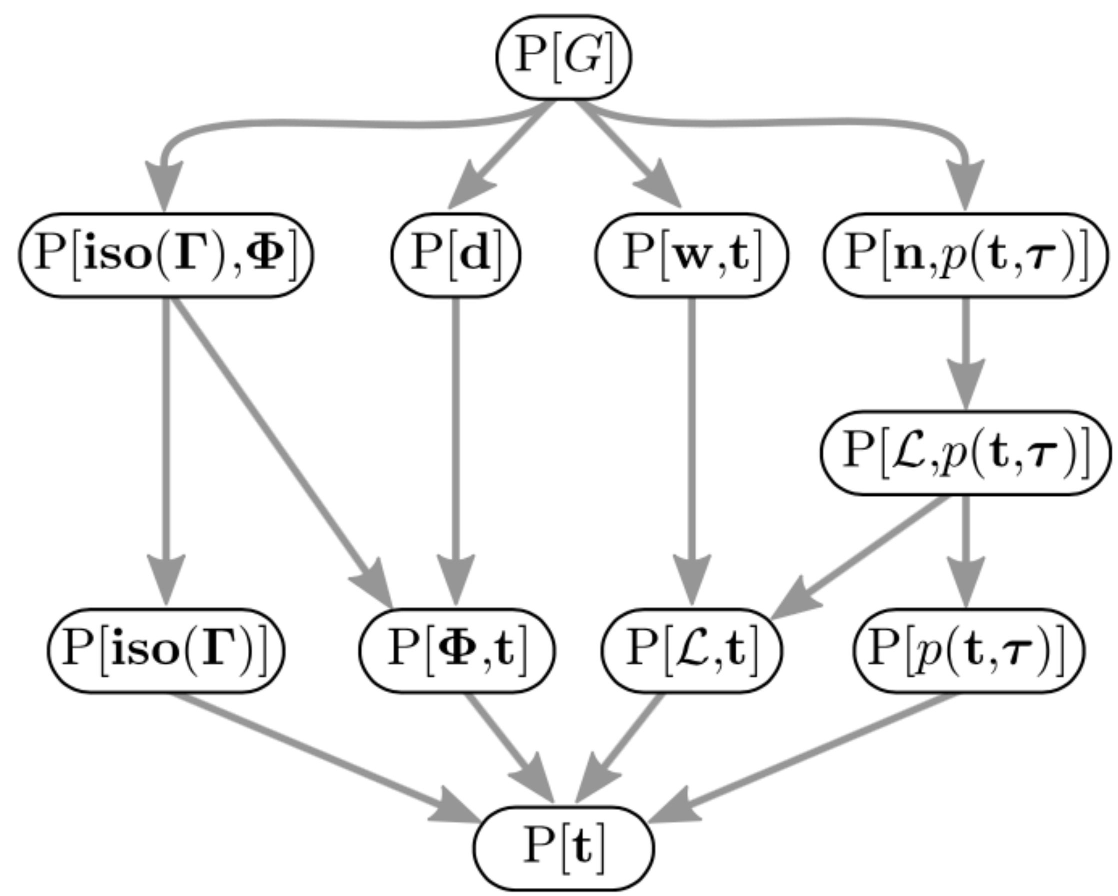


Shuffling methods in the literature

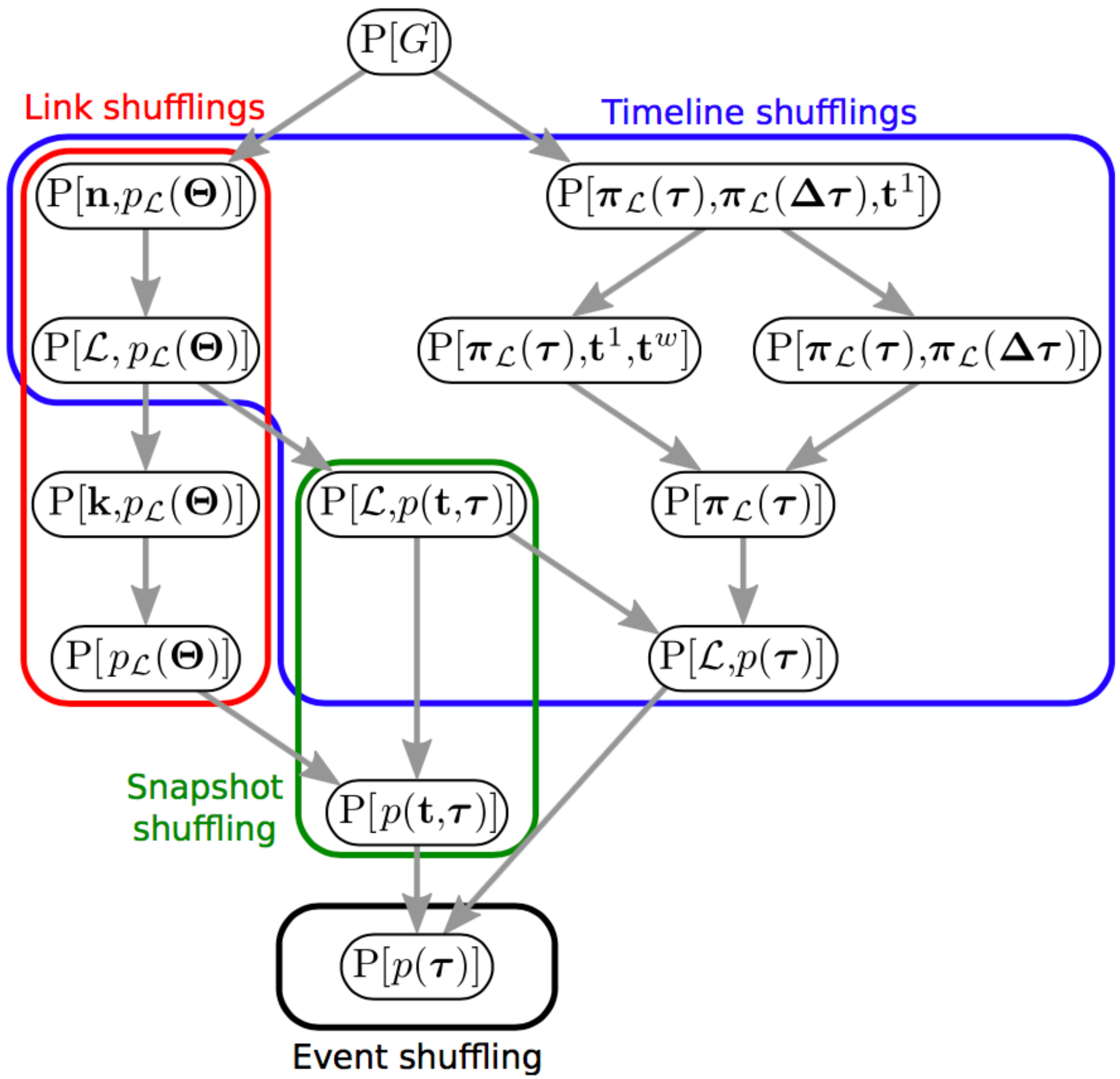
Link shuffling methods



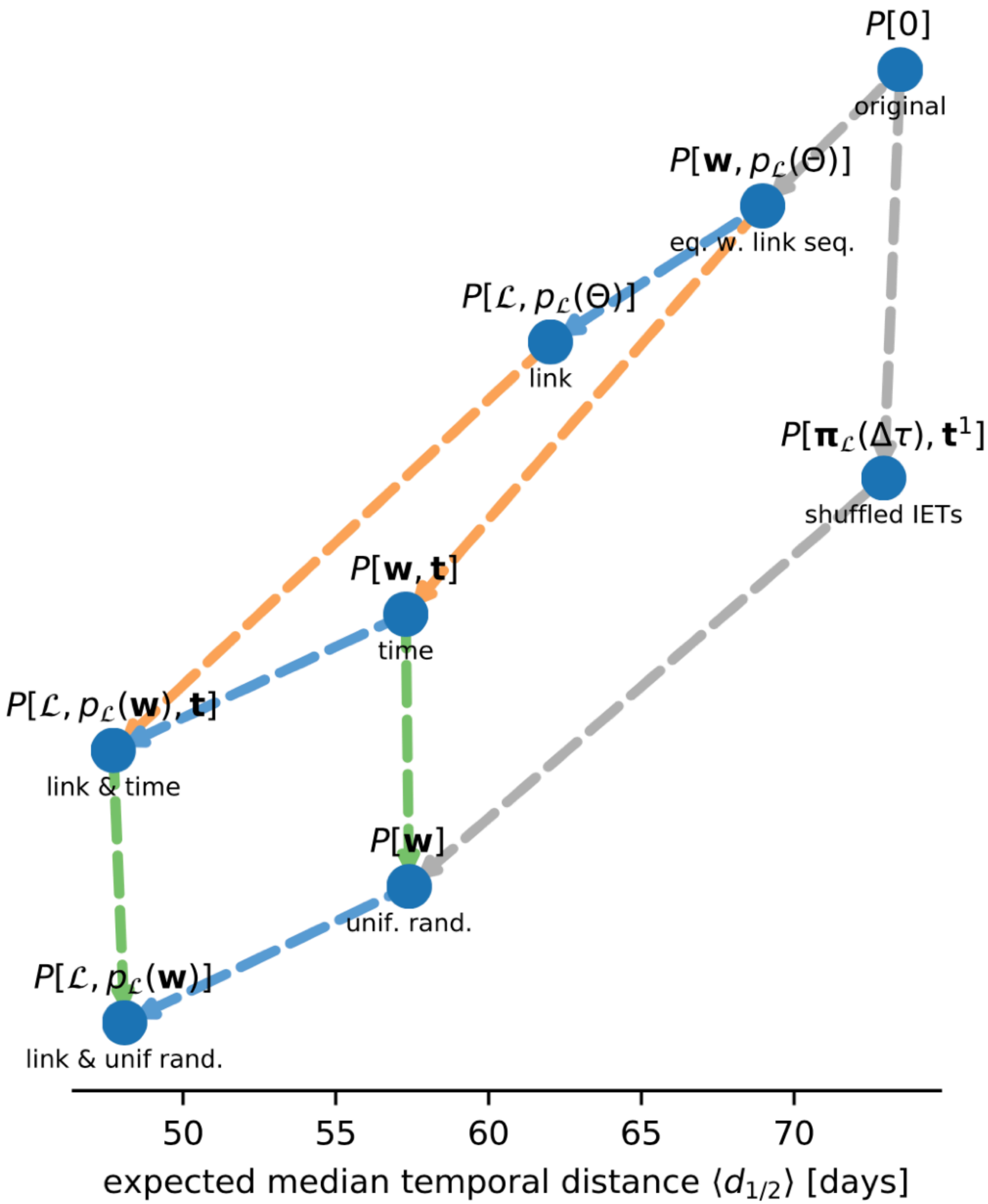
Snapshot shuffling methods



Shuffling methods in the literature



Example analysis



Randomize link activation seqs.
to follow overall activity Θ → **t**
⇒ 12-14 day speedup

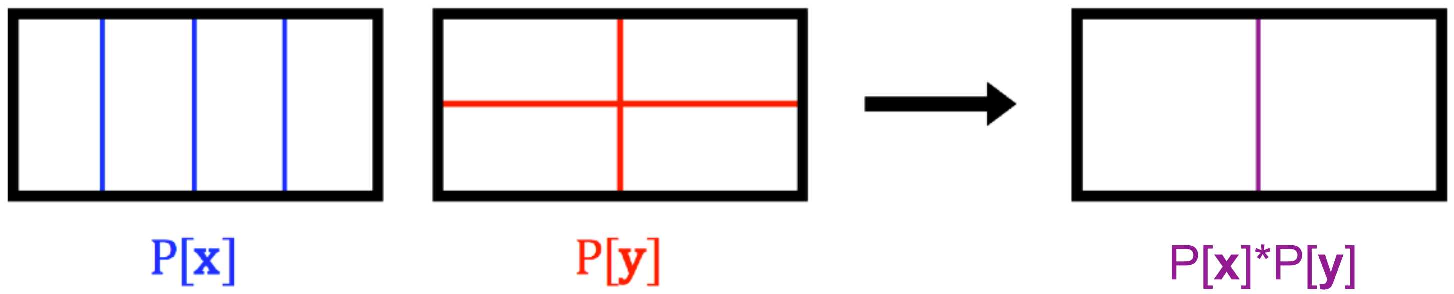
Randomize locations
of link weights **w** →p(**w**)
⇒7-9 day speedup

Randomize overall activity **t**
⇒ no difference

Data from Wu et al. PNAS 107:18803 (2010)

Combining RRMs

- P[**x**]*P[**y**]: a combination, first apply P[**y**] to a network and then apply P[**x**] to the result



- A way of creating new RRMs
- P[**x**]*P[**y**] might not be an MRRM (if they are, we say they are *compatible*)
  - Some clearly compatible cases:
    - P[**x**]*P[**y**] = P[**y**]*P[**x**] = P[**y**] if **x** < **y**
    - "Orthogonality": e.g., P[**x**] permutes time, and P[**y**] permutes topology

Combining MRRMs

- Compatible ⇒ commutative P[**x**]*P[**y**] = P[**y**]*P[**x**]
- General characterisation for compatibility: *Two MRRMs compatible iff they are conditionally independent* given a common coarsening*
- P[**x**], P[**y**] compatible ⇒ P[**x**, g(**y**)] * P[**y**, f(**x**)] = P[**x*****y**, g(**y**), f(**x**)]
- We can create over 100 new MRRMs by combining the ones in the literature

*For detailed definitions etc, see: arXiv:1806.04032

Temporal MRRMs = toolbox for analysing temporal networks

“Randomized reference models for temporal networks” arXiv:1806.04032

Software for MRRMs

- Python/C++ implementation
  - Efficient (handles hundreds of millions of events)
  - *https://github.com/bolozna/Events*
- Pure Python implementation
  - More models
  - *https://github. com/mgenois/RandTempNet*

Some future directions

1. Use the same framework for other higher-order network representations

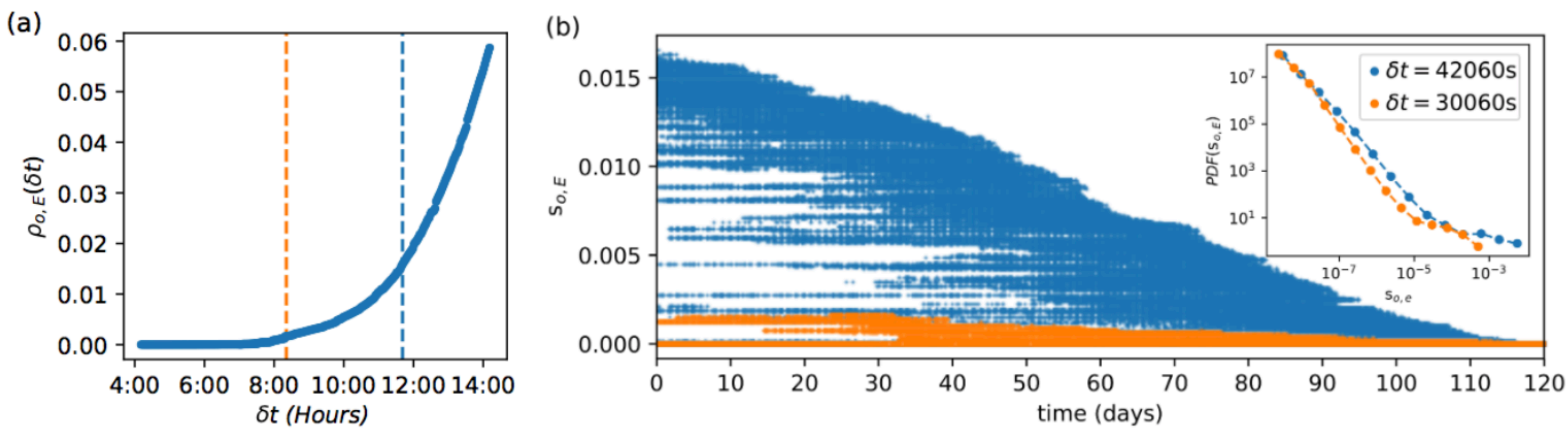
2. Given a subset of constraints, automatically create a shuffling algorithm

3. "Macro-canonical" temporal reference models: *exponential temporal random graphs*

"Randomized reference models for temporal networks" arXiv:1806.04032

Temporal event graphs + probabilistic counting = temporal reachability from all starting points

Fraction of reachable network starting from each of 300M events:



"Efficient limited time reachability estimation in temporal networks", Arash Badie Modiri, M Karsai, MK (in preparation)