Complex networks

Diffusion and spreading on networks

János Török

Department of Theoretical Physics

April 17, 2023

Diffusion on networks

- Random walk
- On lattices we know how it works.
- In what sense will it be different?
- What are the relevant measure for the probability distribution of the walker?
- Why is it important?

Diffusion on one dimensional lattice

- Master equation, lattice and arbitrary coordinates:

$$P(i,t+1) = P(i,t) + \underbrace{\frac{1}{2}P(i-1,t) + \frac{1}{2}P(i+1,t)}_{\text{gain}} - \underbrace{P(i,t)}_{\text{loss}}$$

$$\begin{aligned} P(x,t+\Delta t) =& P(x,t) + \\ & + D\frac{\Delta t}{\Delta x^2}[P(x-\Delta x,t) - 2P(x,t) + P(x+\Delta x,t)] \end{aligned}$$

- Continuum limit: diffusion equation

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}$$

- Solution

$$P(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{x^2}{4Dt}}$$

Diffusion on one dimensional lattice

- Continuum limit: diffusion equation

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}$$

- Solution

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

- Moments of the coordinate

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x,t)dx = 0$$

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 P(x,t)dx = 2Dt$$

Random walk on lattice

- Moments of the coordinate

$$\langle x \rangle = \int_{-\infty}^{\infty} x P(x,t) dx = 0$$
$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 P(x,t) dx = 2Dt$$

- Probability to return to origin (Pólya theorem):

d	p_{ret}
1	1
2	1
3	0.34
4	0.19
5	0.145

Random walk on lattice

- Expected number of distinct sites visited by the random walk

d	D_t
1	$\sim \sqrt{t}$
2	$\sim t/\log t$
$3 \leq d$	$\sim t$

- The trail of the random walk is a fractal with fractal dimension $d = 2$
- In $d = 1$ the trail is self-overlapping
- In $d = 2$ it gradually fills the space
- In $d > 4$ the walk does not cross itself

Random walk on graphs

- Distance is not as important of a quantity as in lattices
- Important quantities:
  - Number of visited distinct sites
  - Probability of return
  - Probability of finding the walker on a given node
  - Probability from going one node to the other

Random walk on Watts-Strogatz graph

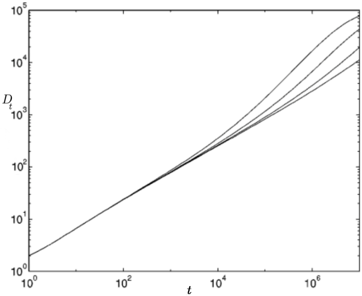
- $p = 0$: We have a one dimensional lattice
- $p = 1$: Random network is similar to trees upon trees, always new regions are explored, or infinite dimension
- Interesting regime $0 < p \ll 1$:
  - Characteristic distance between two crosslink ending: $\xi \sim 1/p$
  - One dimensional system up to $t_\xi \sim \xi^2$
  - Infinite dimension afterwards

Random walk on Watts-Strogatz graph

- Interesting regime $0 < p \ll 1$:
- Characteristic distance between two crosslink ending: $\xi \sim 1/p$
- One dimensional system up to $t_\xi \sim \xi^2$
- Infinite dimension afterwards
- Number of visited distinct sites:

$$D_t = \sqrt{t} f(t/t\xi) = \sqrt{t} f(tp^2)$$

$$f(x) = \begin{cases} \text{const} & \text{if } x \ll 1 \\ \sqrt{x} & \text{if } x \gg 1 \end{cases}$$



Random walk on graphs

- Let r be the rate of leaving a site
- The walker at node i
- Moves randomly to any neighbour, with the same probability
- Nodes are characterized by their degree k_i
- In order to land on a node with degree k from a node with degree k' the latter must have a neighbour with degree k
- The probability of going from a node with degree k' to a node with degree k is $P(k'|k)/k'$, where the former is the probability of a node with degree k' have a neighbour with degree k (assortativity)
- Master equation ($n_k(t)$ number of walkers on nodes with degree k)

$$\frac{\partial n_k(t)}{\partial t} = -rn_k(t) + rk\sum_{k'} P(k'|k)n_{k'}(t)/k'$$

Random walk on graphs

- Master equation ($n_k(t)$ number of walkers on nodes with degree k)

$$\frac{\partial n_k(t)}{\partial t} = -r n_k(t) + rk \sum_{k'} P(k'|k) n_{k'}(t)/k'$$

- The first term is the loss term: walkers leave with rate r
- The gain term is proportional to
  - Walking rate
  - The degree of the node k (walkers may come in through k links)
  - The probability that it comes from a node with degree k'

Random walk on graphs

- Master equation ($n_k(t)$ number of walkers on nodes with degree k)

$$\frac{\partial n_k(t)}{\partial t} = -r n_k(t) + rk \sum_{k'} P(k'|k) n_{k'}(t)/k'$$

- For uncorrelated networks we have

$$P(k'|k) = \frac{k' P(k')}{\langle k \rangle}$$

- Which leads to

$$\frac{\partial n_k(t)}{\partial t} = -r n_k(t) + r \frac{k}{\langle k \rangle} \sum_{k'} n_{k'}(t)$$

Random walk on graphs

- Master equation on uncorrelated graphs

$$\frac{\partial n_k(t)}{\partial t} = -rn_k(t) + r\frac{k}{\langle k \rangle}\sum_{k'} n_{k'}(t)$$

- The stationary solution (left hand side vanishes):

$$n_k = \frac{k}{\langle k \rangle}\frac{n}{N},$$

where n is the number of walkers. Or with probability

$$p_k = \frac{k}{\langle k \rangle}\frac{1}{N},$$

where p_k is the probability of finding the walker at a node with degree k

Random walk on graphs

- The probability of finding the walker at a node with degree k

$$p_k = \frac{k}{\langle k \rangle} \frac{1}{N},$$

- It is more likely to find the walkers at hubs than in a dead end
- *There are more drunk people at Deák tér and at Nyugati than e.g. at Gárdonyi tér.*

Diffusion equation on graphs

- Recall diffusion equation on 1d lattice:

$$\Phi(x, t+\Delta t) = \Phi(x, t)+D\Delta t[\Phi(x-\Delta x, t)-2\Phi(x, t)+\Phi(x+\Delta x, t)]$$

- Which can be rewritten as

$$\Phi(x, t + \Delta t) = \Phi(x, t) + dtD\mathrm{L}\Phi(x, t),$$

where

$$\mathrm{L}\Phi(x, t) = \sum_{dx\in\pm\Delta x} \Phi(x + dx) - \Phi(x) \sum_{dx\in\pm\Delta x} 1$$

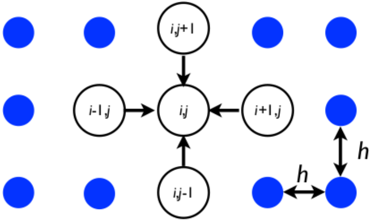
- Multiple dimensions:

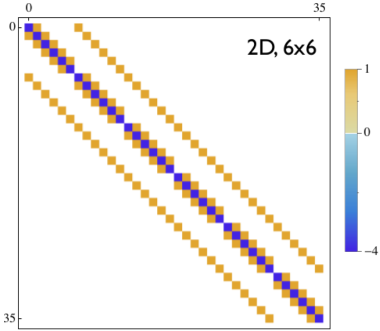
$$\mathrm{L}\Phi(\mathrm{r}, t) = \sum_{d\mathrm{r}\in nn.} \Phi(\mathrm{r} + d\mathrm{r}) - \Phi(\mathrm{r}) \sum_{d\mathrm{r}\in nn.} 1$$

Diffusion equation on graphs

- Diffusion equation on lattices

$$L\Phi(\mathrm{r}, t) = \sum_{d\mathrm{r} \in nn.} \Phi(\mathrm{r} + d\mathrm{r}) - \Phi(\mathrm{r}) \sum_{d\mathrm{r} \in nn.} 1$$



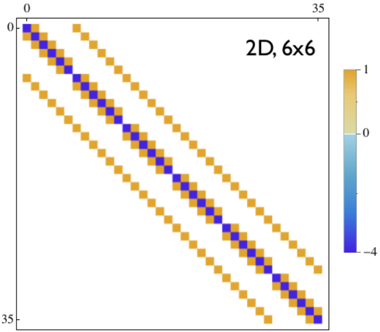


Diffusion equation on graphs

- Diffusion equation on lattices

$$\mathrm{L}\Phi(\mathrm{r}, t) = \sum_{d\mathrm{r} \in nn.} \Phi(\mathrm{r} + d\mathrm{r}) - \Phi(\mathrm{r}) \sum_{d\mathrm{r} \in nn.} 1$$

- Laplace matrix has 1 values where the adjacency matrix would also be 1 and apart from the diagonal is zero where the adjacency matrix would be 0
- The diagonal is minus the degree of the node.



Diffusion equation on graphs

- Diffusion equation on lattices

$$L\Phi(r, t) = \sum_{dr \in nn.} \Phi(r + dr) - \Phi(r) \sum_{dr \in nn.} 1$$

- Generalization to graphs

$$L_{ij} = A_{ij} - k_i \delta_{ij}$$

- Valid also for directed graphs:
  - Not symmetric
  - In diagonal k_i^{out}

Spectral analysis

- Diffusion operator on graphs

$$L_{ij} = A_{ij} - k_i \delta_{ij}$$

- Spectral analysis

$$\sum_j L_{ij} u_j = \lambda_i u_i$$

- Larges eigenvalue: 0, Eigenvector: $(1, 1, 1, \dots)$ with multiplicity equals to the number of connected components
- Second largest eigenvalue shows how difficult it is to split the graph into two large pieces. (How easy it is to reach all parts of the network)

$$\begin{array}{ll} \lambda^{(2)} = -n & \text{for an } n\text{-clique} \\ \lambda^{(2)} = -1 & \text{for a star} \\ \lambda^{(2)} = -2 + 2\cos(\pi/n) & \text{for an } n\text{-chain} \end{array}$$

- The last one goes to zero for $n \to \infty$

Spectral analysis of the diffusion operator

- Diffusion equation on graph
- Eigenvalue distribution (average them over all node):

$$\rho(\lambda) = \left\langle \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda^{(i)}) \right\rangle$$

- Initial condition: walker on node i_0 at $t{=}0$
- Probability to be at node i at time t

$$\frac{\partial p(i,t|i_0,0)}{\partial t} = \sum_j L_{ij} p(j,t|i_0,0)$$

- Laplace transform:

$$\tilde{p}_{i,i_0}(s) = \int_0^\infty e^{-st} p(i,t|i_0,0) dt$$

Spectral analysis of the diffusion operator

- Diffusion equation on graph

$$\frac{\partial p(i,t|i_0,0)}{\partial t} = \sum_j L_{ij} p(j,t|i_0,0)$$

- Laplace transform:

$$\tilde{p}_{i,i_0}(s) = \int_0^\infty e^{-st} p(i,t|i_0,0) dt$$

- From the diffusion equation

$$s\tilde{p}_{i,i_0} - \delta_{i,i_0} = \sum_j L_{ij} \tilde{p}_{j,i_0}$$

- $f'(t) \quad \rightarrow$ Laplace transform $\rightarrow \quad sF(s) - f(0^+)$

Spectral analysis of the diffusion operator

- Laplace transform:

$$\tilde{p}_{i,i_0}(s) = \int_0^\infty e^{-st} p(i,t|i_0,0)dt$$

- From the diffusion equation

$$s\tilde{p}_{i,i0} - \delta_{i,i0} = \sum_j L_{ij}\tilde{p}_{j,i_0}$$

- From where

$$\sum_j (s\delta_{i,j} - L_{ij})\tilde{p}_{j,i_0} = \delta_{i,i_0}$$

Spectral analysis of the diffusion operator

- Probability to return to the origin

$$p_0(t) = \left\langle \frac{1}{N} \sum_{i_0} p(i_0, t | i_0, 0) \right\rangle$$

- Laplace transform

$$\tilde{p}_0(s) = \left\langle \frac{1}{N} \sum_{i_0} \tilde{p}(i_0, t | i_0, 0) \right\rangle = \left\langle \frac{1}{N} \mathrm{Tr} \tilde{p}(i_0, t | i_0, 0) \right\rangle =$$

$$= \left\langle \frac{1}{N} \mathrm{Tr} \left(s\delta_{ij} - L_{ij} \right)^{-1} \right\rangle = \left\langle \frac{1}{N} \sum_{i} \frac{1}{s - \lambda^{(i)}} \right\rangle$$

Spectral analysis of the diffusion operator

- Probability to return to the origin

$$p_0(t) = \left\langle \frac{1}{N} \sum_{i_0} p(i_0, t | i_0, 0) \right\rangle$$

- Laplace transform

$$\tilde{p}_0(s) = \left\langle \frac{1}{N} \sum_{i_0} \tilde{p}(i_0, t | i_0, 0) \right\rangle = \left\langle \frac{1}{N} \sum_{i} \frac{1}{s - \lambda^{(i)}} \right\rangle$$

- Transfer back

$$p_0(t) = \int e^{ts} \left\langle \frac{1}{s - \lambda^{(i)}} \right\rangle ds = \left\langle \frac{1}{N} \sum_{i} e^{\lambda^{(i)} t} \right\rangle$$

$$p_0(t) = \int_{-\infty}^{0} e^{t\lambda} \rho(\lambda) d\lambda$$

Spectral analysis of the diffusion operator

- Probability to return to the origin

$$p_0(t) = \int_{-\infty}^{0} e^{t\lambda}\rho(\lambda)d\lambda$$

- The shape of the spectrum thus determines the return probability
- Example: Watts-Strogatz small world

$$p_0(t) - p_0(0) \sim \begin{cases} t^{-d/2} & \text{if } t \ll t_\xi \\ \exp\left(-(p^2 t)^{1/3}\right) & \text{if } t \gg t_\xi \end{cases}$$

- The spectrum of the Laplacian is related also to the community structure of the network
- The largest eigenvalue describes the stationary state.
- The second largest is related to processes longest time scales.

Transition probability

- Transition probability from node i to j.
- We can exit node i an any of its link
- We can enter node j only of there is a connection

$$P_{ij} = \frac{A_{ij}}{k_i}$$

- The probability of going from i to j in t steps is:

$$P_{i \to j}(t) = \sum_k P_{ik} \sum_l P_{kl} \sum_m P_{lm} \cdots \sum_v P_{sv} P_{vj} = \left(P^t\right)_{ij}$$

- P^t is the tth power of the P matrix
- Distance measure

$$r_{ij}(t) = \sqrt{\sum_{l=1}^{N} \frac{(P^t_{il} - P^t_{jl})^2}{k_l}}$$

Temporal networks

- Links are not always present
- Examples:
  - Communication networks
  - Public transportation
  - Company contracts/orders
  - Spreading
  - Time evolution of the network
- If timescales separate we can study temporal events over a static networks
- Aggregate network: all links and nodes ever present

MOVIE

Peter Holme - Jari Saramäki (2011) Arxiv:1108.1780

Network definition

- Static network: $G = \{V, E\}$
- Temporal network: $T = \{V, S\}$, where V is the set of vertices and S is the set of event sequences (can be directed)
- For $s_{ij} \in S$
$$s_{ij} = \left\{ t_{ij}^{(1)}, \tau_{ij}^{(1)}; t_{ij}^{(2)}, \tau_{ij}^{(2)}; \ldots \right\}$$
- where event r between node i and j begins at $t_{ij}^{(r)}$ and lasts $\tau_{ij}^{(r)}$
- $\tau_{ij}^{(r)}$ can often be neglected
- Adjacency index
$$A(i,j,t) = \begin{cases} 1 & \text{if } i \to j \text{ is active at time } t \\ 0 & \text{otherwise} \end{cases}$$

Adjacency index

- Adjacency index

$$A(i,j,t) = \begin{cases} 1 & \text{if } i \to j \text{ is active at time } t \\ 0 & \text{otherwise} \end{cases}$$

- Adjacency index for instantaneous events

$$A(i,j,t,\Delta t) = \begin{cases} 1 & \text{if } i \to j \text{ is active between time } t \text{ and } t + \Delta T \\ 0 & \text{otherwise} \end{cases}$$

- Conditional aggregate networks: $A(i,j,t,\Delta t)$

Temporal networks: path, journey

- **Path**: series of distinct edges visiting distinct nodes
- **Journey**: a time respecting path, time window (t_{min}, t_{max})

$$J_{1\to n} = \{t_{12}, t_{23}, \dots, t_{n-1,n} | t_{ij} \in S, t_{min} \leq t_{12} \leq \dots \leq t_{n-1,n} \leq t_{max}\}$$

- **Reachibility**: i is reachable from j, if there exists a journey from i to j
- **Set of influence**: all nodes which are reachable from i

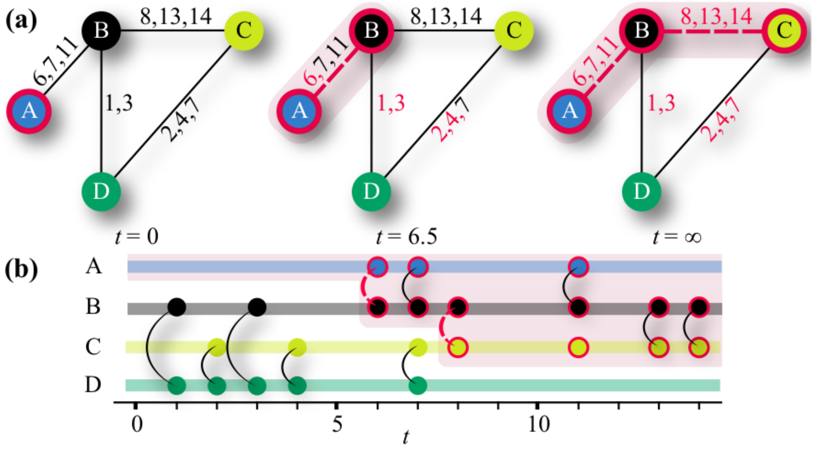
$$I_i(t) = \{\forall j | j \in V, \exists J_{i\to j}\}$$

- **Source set**: all nodes from which i is reachable

$$S_i(t) = \{\forall j | j \in V, \exists J_{j\to i}\}$$

Temporal networks: visualization

- Journeys are non-transitive: $\exists J_{A\to B}$ and $\exists J_{B\to C}$, but $\nexists J_{A\to C}$
- $I_A = \{B, C\}$, $S_A = \{B, C, D\}$
- $I_C(t \in [5, 10]) = \{B, D\}$, $S_C(t \in [5, 10]) = \{A, B, D\}$



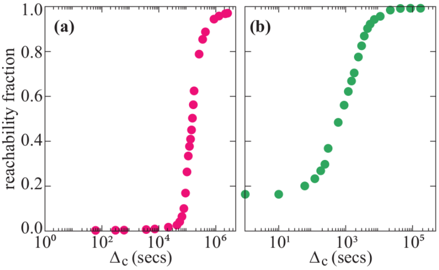
Temporal networks: reachability

- **Journeys with maximal waiting times**: a time respecting path, with limited event separation

$$J^{\Delta t}_{1\to n} = \{t_{12}, \ldots, t_{n-1,n} | t_{ij} \in S, t_{12} \leq \cdots \leq t_{n-1,n}; t_{i+1} - t_i < \Delta t\}$$

- Reachability ratio: average fraction of nodes reachable from each node

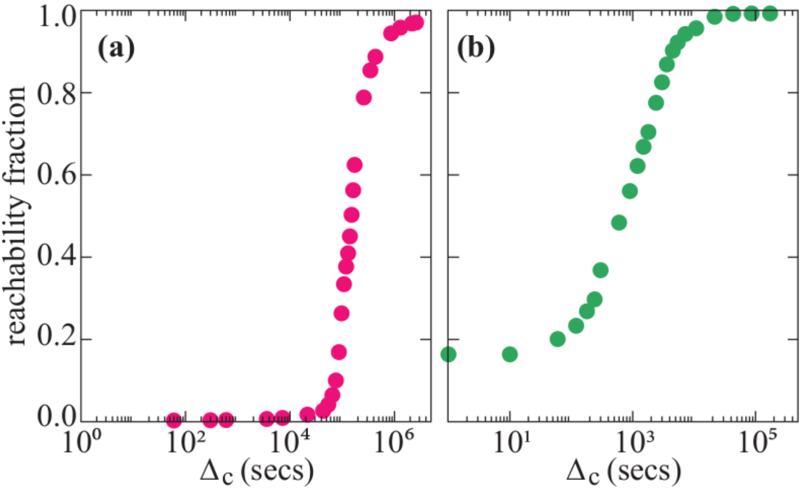
$$r^{\Delta t}(t) = \frac{1}{N} \sum_i |I_i^{\Delta t}(t)|$$



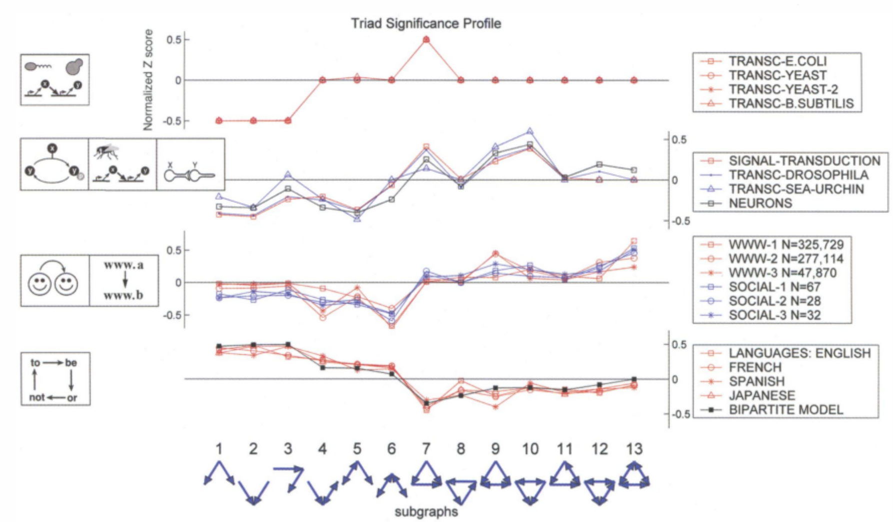
phone call airline

Temporal networks: reachability

phone call　　　　airline

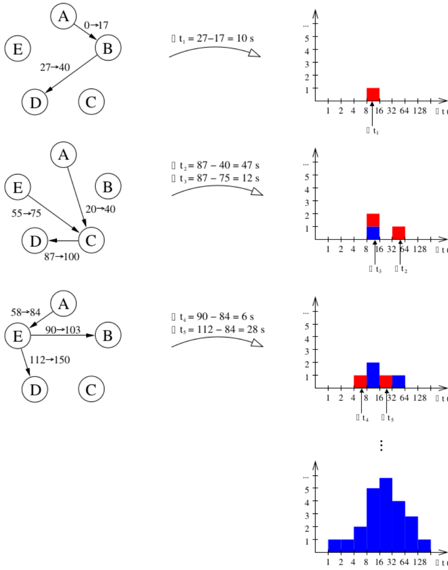


Static motifs



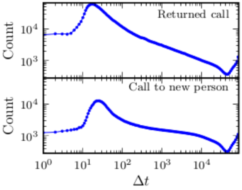
Action triggers

- Detect causal chains of events
- Measure typical reaction time
- Measure waiting time between incoming and outgoing calls
- Make histogram from it

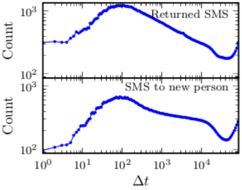


Action triggers histogram

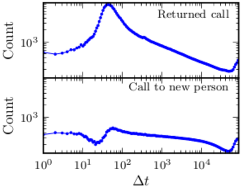
- Maximum occurs at 17 seconds for returned calls
- Maximum occurs at 25 seconds for calls to a new person
- SMS peaks are typically 20-24 seconds later
- You need that much time to read and write an SMS



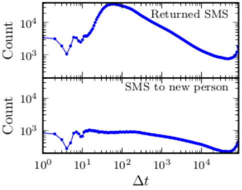
(a) From call to call



(b) From call to SMS



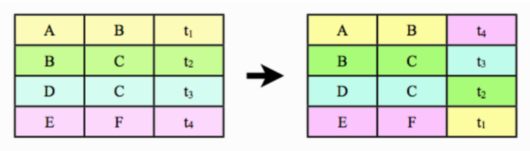
(c) From SMS to call



(d) From SMS to SMS

Temporal motifs

- Now we know the relevant timescales
- We detect topological objects within the defined time window
- Sliding window over the whole data
- Null model: Shuffled time reference

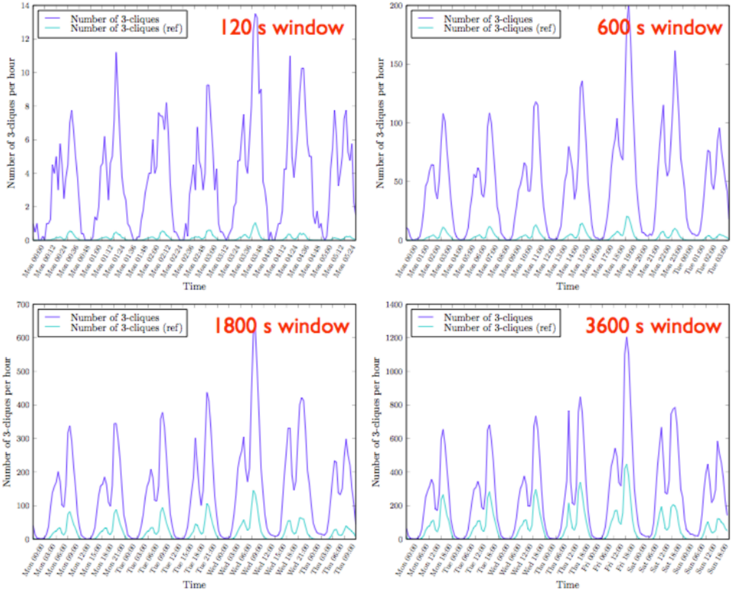


A	B	t_1
B	C	t_2
D	C	t_3
E	F	t_4

→

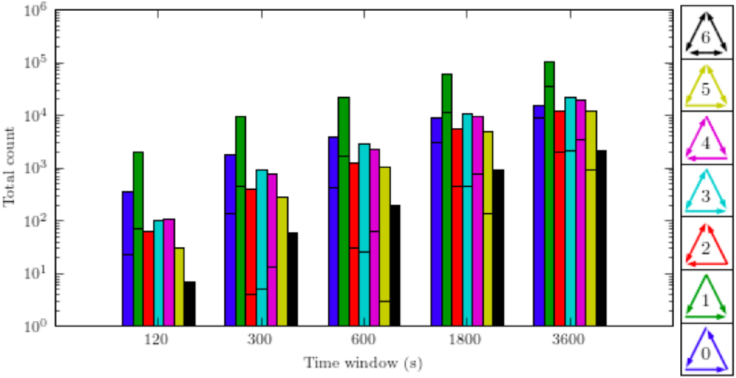
A	B	t_4
B	C	t_3
D	C	t_2
E	F	t_1

Temporal motifs: occurrence of triangles



Temporal motifs: occurrence of triangles

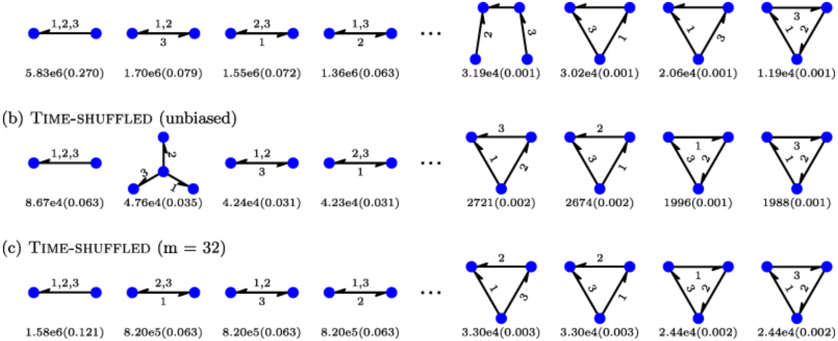
- Without order
- Horizontal line: time shuffled reference



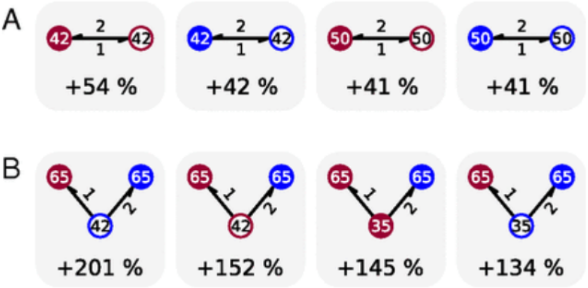
Temporal motifs: occurrence of ordered sequences

Most frequent ones

Least frequent ones



Example of temporal effects



Dyn.
1st pre to post

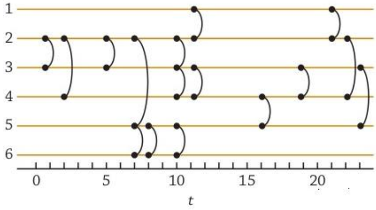
Outstar: Target
age same

Female, 42 ± 2 years old, prepaid user

Male, 50 ± 2 years old, postpaid user

Spreading on temporal networks

- Links are not always present
- This definitely slows down the spreading
- This effect can be considerable



Importance of different effects in temporal spreading

- Original data: time ordered sequence of call events
- It contains information about the underlying network
- Correlations:
  - D: daily pattern
  - C: community structure
  - W: weight-topology
  - B: bursty single-edge dynamics
  - E: event-event

Karsai et al. PRE 2011

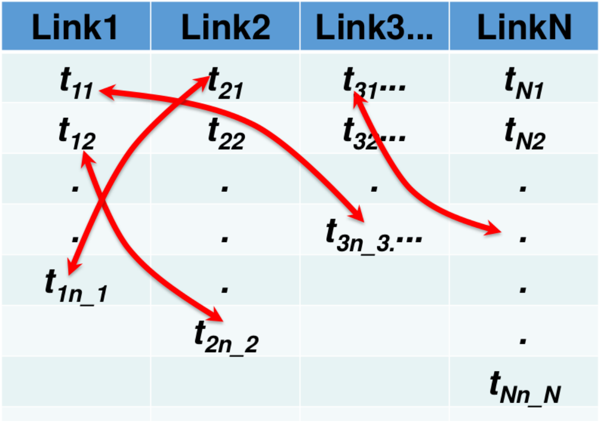
Link shuffling

- Select random pairs of link sequences and exchange them
- Destroys topology-weight and link-link correlation

Link1	Link2	Link3...	LinkN
t_{11}	t_{21}	t_{31}...	t_{N1}
t_{12}	t_{22}	t_{32}...	t_{N2}
.	.	.	.
.	.	t_{3n_3}...	.
t_{1n_1}	.		.
	t_{2n_2}		.
			t_{Nn_N}

Time shuffling

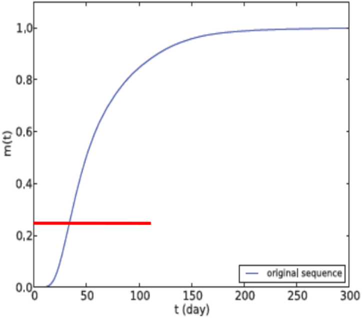
- Destroys burstiness (and link-link correlations)
- Keeps weight and daily pattern



Importance of different effects in temporal spreading

- Original data: time ordered sequence of call events

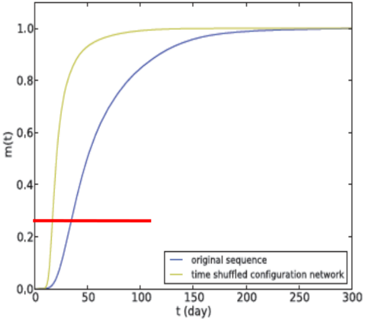
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7



Importance of different effects in temporal spreading

- Configuration model: Network is rewired, community structure destroyed
- Event times are shuffled: Bursty dynamics destroyed

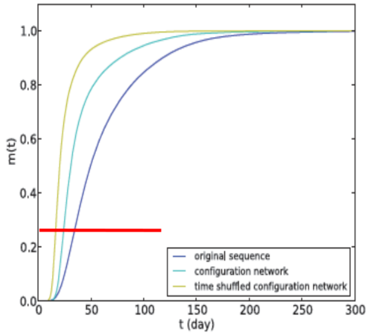
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. model	✓	×	×	×	×	16.4



Importance of different effects in temporal spreading

- Configuration model: Network is rewired, community structure destroyed
- Event times are kept Bursty dynamics kept

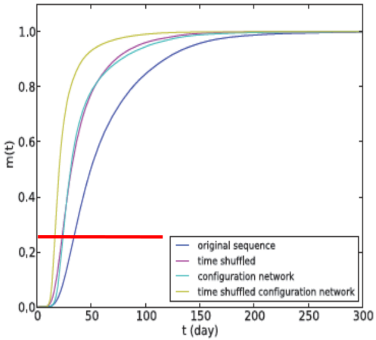
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8



Importance of different effects in temporal spreading

- Time shuffled event sequence
- Bursty dynamics destroyed
- Community and weight topology correlations kept

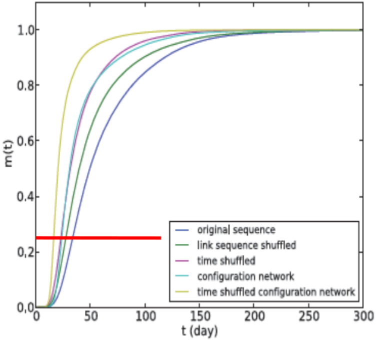
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9



Importance of different effects in temporal spreading

- Link sequence shuffled
- Link-link and weight topology is destroyed
- Bursty dynamics and community structure is kept

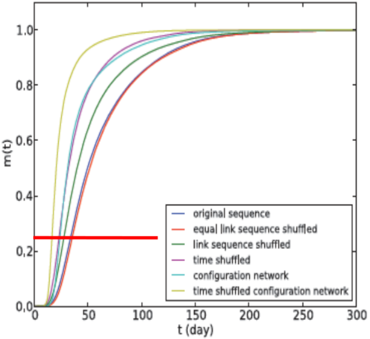
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9
Shuffle. keep	✓	✓	×	✓	×	27.5



Importance of different effects in temporal spreading

- Equal-weight link-sequence shuffled: Whole single-link event sequences are randomly exchanged between links having the same number of events
- Only link-link correlation is destroyed

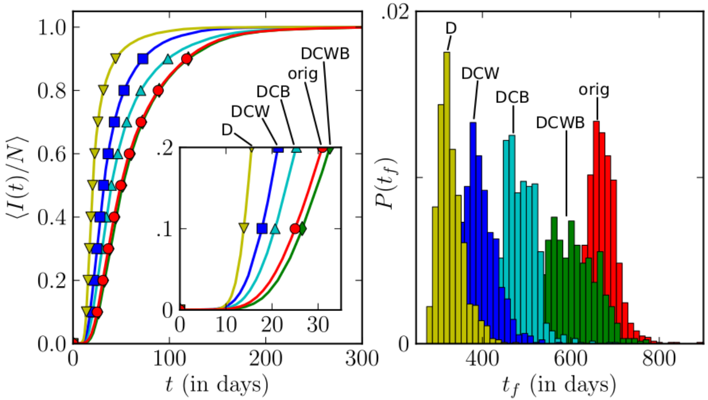
Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9
Shuffle. keep	✓	✓	×	✓	×	27.5
W keep sh.,keep	✓	✓	✓	✓	×	35.3



Importance of different effects in temporal spreading

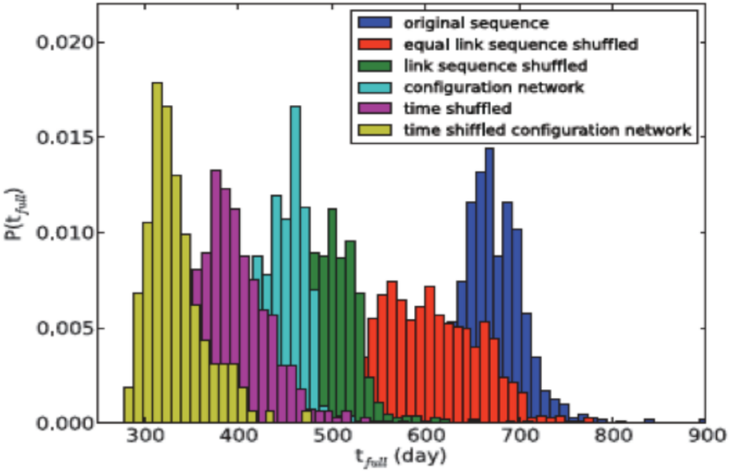
- Long time behaviour:

Event sequence	D	C	W	B	E	25%
Original	✓	✓	✓	✓	✓	33.7
Config. shuffle	✓	×	×	×	×	16.4
Config. keep	✓	×	×	✓	×	23.8
Orig. shuffle	✓	✓	✓	×	×	22.9
Shuffle. keep	✓	✓	×	✓	×	27.5
W keep sh.,keep	✓	✓	✓	✓	×	35.3



Importance of different effects in temporal spreading

- Everything slows down the spreading
- Burstiness has higher impact than topological structures

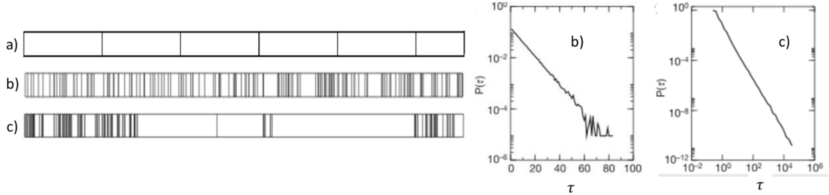


Interevent time

- Time interval between successive events τ
- Distribution of τ is $P(\tau)$
- Distribution is characterized by the average $\langle\tau\rangle$ and the variance σ
- Burstiness:

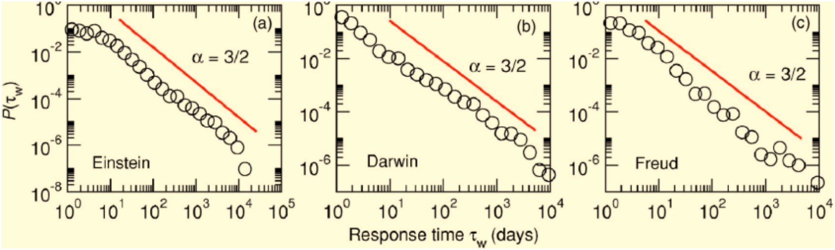
$$B = \frac{\sigma - \langle\tau\rangle}{\sigma + \langle\tau\rangle}$$

- (a) $B = -1$: deterministic, (b) $B = 0$: Poisson, (c) $B = 1$: bursty



Bursty examples:

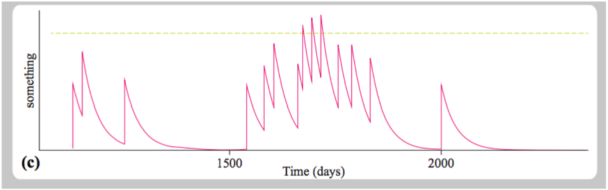
- Response times for letters



A. Vázquez PRE 2006

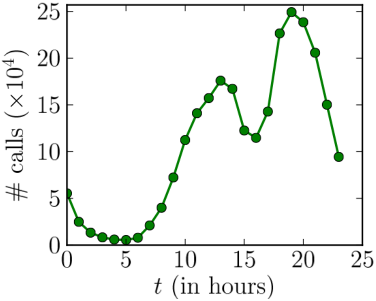
Reason of bursty behavior

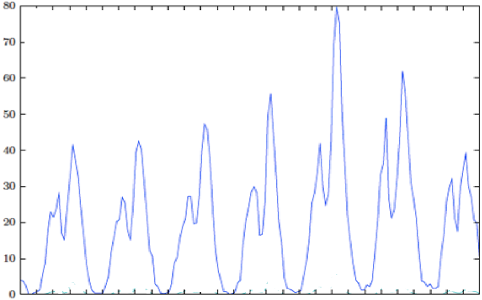
- Highly concentrated events
- If you pick up phone you complete more tasks
- If an old friend called you it is more probable that you call him back soon



Seasoning

- Problem with day/week/year





Deseasoning

- Rescedule the events to be periodic over a period T
- Let i be an individual
- $n_i(t) = 1$ if there is an event $n_i(t) = 0$ if there is not

$$s_i(t) = \sum_{t'=0}^{t} n_i(t')$$

- Strength of node i over the observation period
- For a set of people Λ, the number of events at time t

$$n_\Lambda(t) = \sum_{i \in \Lambda} n_i(t)$$

Jo et al., Circadian pattern and burstiness in mobile phone communication (2011)

Deseasoning

- Rescaled event rate

$$\rho_{\Lambda,T}(t) = \frac{T}{s_\Lambda} \sum_{k=0}^{T_f/T} n_\Lambda(t+kT) \qquad s_\Lambda = \sum_{t=0}^{T_f} n_\Lambda(t)$$

$$\tau^* = t^*(t_{j+1}) - t^*(t_j) = \sum_{0 \le t' < t} \rho_{\Lambda,T}(t')$$