Modeling sequences and temporal networks with dynamic community structures

Tiago P. Peixoto

Universität Bremen & ISI Foundation

Martin Rosvall

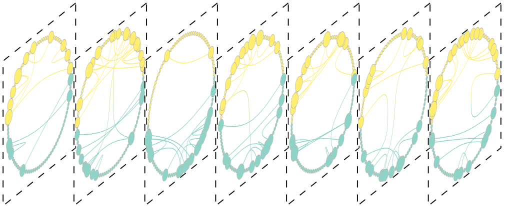
Umeå University, Sweden

Seoul, May 2016

How to model dynamic network structure?

Common idea: Temporal coarsening

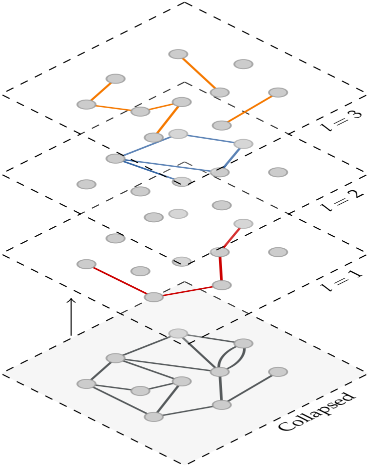
Adapt a *static* network approach into a dynamical one via aggregation into *temporal layers*.



⟶ Time

Stochastic Block Models with layers

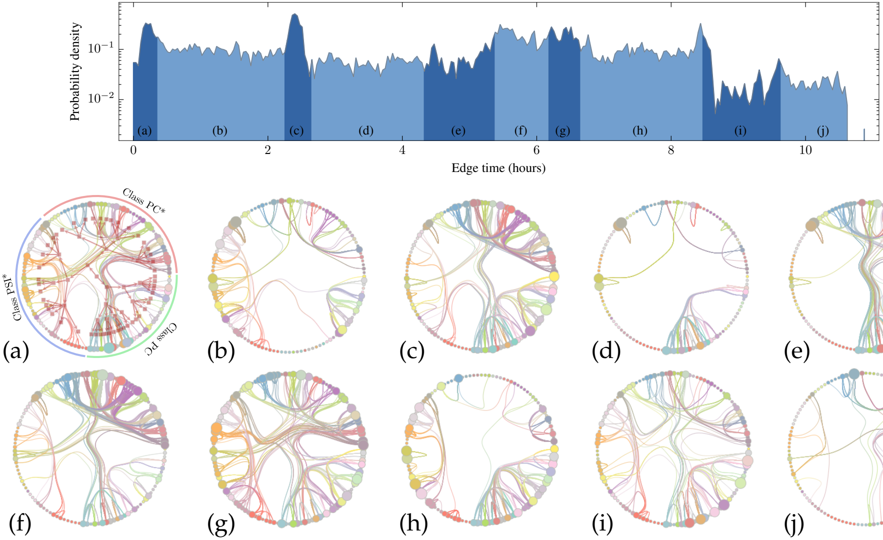
T.P.P, Phys. Rev. E 92, 042807 (2015)



- Fairly straightforward. Easily combined with degree-correction, overlaps, etc.
- Edge probabilities are in general different in each layer.
- Node memberships can move or stay the same across layer.
- Works as a general model for discrete as well as *discretized* edge covariates.
- Works as a model for temporal networks.

Proximity between high-school students

T.P.P, Phys. Rev. E 92, 042807 (2015)



J. Fournet, A. Barrat. PLoS ONE 9(9):e107878 (2014), `http://www.sociopatterns.org/`

Some problems with layers

- Time bins must be chosen (a priori or a posteriori [1]).
- Dynamics at a shorter time scale is obscured.
- Statistical independence between bins $\rightarrow$ Evolutionary dynamics is not learned.
- Weak predictive power (usually depends on the very last layer).
- Not very dynamic...

[1] T.P.P Phys. Rev. E 92, 042807 (2015)

Temporal networks redux

As an actual dynamical system this time

A more basic scenario: Sequences.

$$x_t \in [1, N] \rightarrow \text{ Token at time } t$$

$$x_t = \{x_0, x_1, x_2, x_3, \dots\}$$

(Tokens could be letters, base pairs, itinerary stops, *edges of a network*, etc.)

n-ORDER MARKOV CHAINS WITH COMMUNITIES

T. P. P. AND MARTIN ROSVALL, ARXIV: 1509.04740

Transitions conditioned on the last n tokens

$p(x_t|\vec{x}_{t-1}) \rightarrow$ Probability of transition from memory $\vec{x}_{t-1} = \{x_{t-n}, \ldots, x_{t-1}\}$ to token x_t

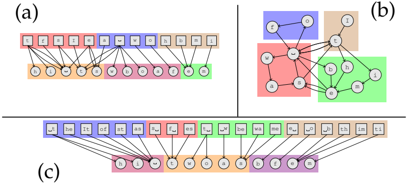
Instead of such a direct parametrization, we divide the tokens and memories into groups:

$$p(x|\vec{x}) = \theta_x \lambda_{b_x b_{\vec{x}}}$$

$\theta_x \rightarrow$ Overall frequency of token x

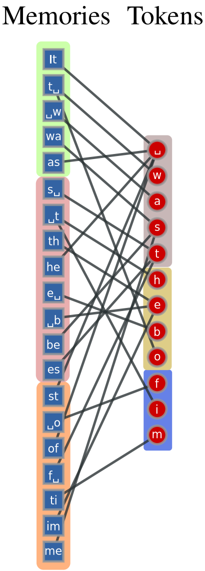
$\lambda_{rs} \rightarrow$ Transition probability from memory group s to token group r

$b_x, b_{\vec{x}} \rightarrow$ Group memberships of tokens and groups



$\{x_t\}$ = "It␣was␣the␣best␣of␣times"

n-ORDER MARKOV CHAINS WITH COMMUNITIES



$$\{x_t\} = \texttt{"It␣was␣the␣best␣of␣times"}$$

$$P(\{x_t\}|b) = \int d\lambda d\theta\, P(\{x_t\}|b,\lambda,\theta)P(\theta)P(\lambda)$$

The Markov chain likelihood is (almost) identical to the SBM likelihood that generates the bipartite transition graph.

Nonparametric $\rightarrow$ We can select the **number of groups** *and* the **Markov order** based on statistical evidence!

T. P. P. and Martin Rosvall, arXiv: 1509.04740

Bayesian formulation

$$P(\{x_t\}|b) = \int \mathrm{d}\theta\, \mathrm{d}\lambda\, P(\{x_t\}|b,\lambda,\theta) \prod_r \mathcal{D}_r(\{\theta_x\}) \prod_s \mathcal{D}_s(\{\lambda_{rs}\})$$

Noninformative priors → Microcanonical model

$$P(\{x_t\}|b) = P(\{x_t\}|b,\{e_{rs}\},\{k_x\}) \times P(\{k_x\}|\{e_{rs}\},b) \times P(\{e_{rs}\}),$$

where

$P(\{x_t\}|b,\{e_{rs}\},\{k_x\}) \to$ Sequence likelihood,

$P(\{k_x\}|\{e_{rs}\},b) \to$ Token frequency likelihood,

$P(\{e_{rs}\}) \to$ Transition count likelihood,

$-\ln P(\{x_t\},b) \to$ ***Description length*** of the sequence

Inference ↔ Compression

n-ORDER MARKOV CHAINS WITH COMMUNITIES

	US Air Flights				War and peace				Taxi movements				"Rock you" password list			
n	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'	B_N	B_M	Σ	Σ'
1	384	365	364,385,780	365,211,460	65	71	11,422,564	11,438,753	387	385	2,635,789	2,975,299	140	147	1,060,272,230	1,060,385,582
2	386	7605	319,851,871	326,511,545	62	435	9,175,833	9,370,379	397	1127	2,554,662	3,258,586	109	1597	984,697,401	987,185,890
3	183	2455	318,380,106	339,898,057	70	1366	7,609,366	8,493,211	393	1036	2,590,811	3,258,586	114	4703	910,330,062	930,926,370
4	292	1558	318,842,968	337,988,629	72	1150	7,574,332	9,282,611	397	1071	2,628,813	3,258,586	114	5856	889,006,060	940,991,463
5	297	1573	335,874,766	338,442,011	71	882	10,181,047	10,992,795	395	1095	2,664,990	3,258,586	99	6430	1,000,410,410	1,005,057,233
gzip	573,452,240				9,594,000				4,289,888				1,315,388,208			
LZMA	402,125,144				7,420,464				2,902,904				1,097,012,288			

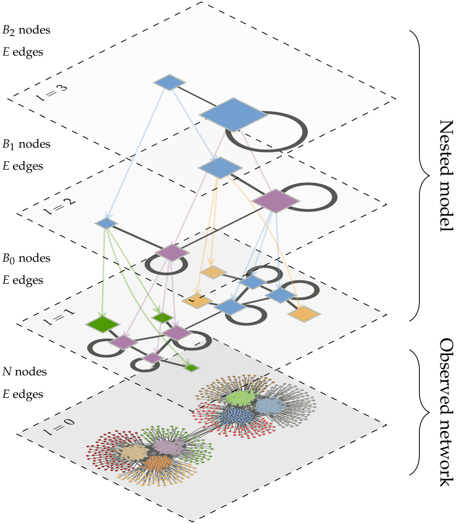
(SBM can compress your files!)

T. P. P. and Martin Rosvall, arXiv: 1509.04740

CHOOSING PRIORS: BAYESIAN HIERARCHIES

T. P. PEIXOTO, PHYS. REV. X 4, 011047 (2014)

Problem: Flat priors can *underfit*, $B_{\max} \sim \sqrt{N}$ [1]



Higher resolution

$$B_{\max} \sim \frac{N}{\ln N} \gg \sqrt{N}$$

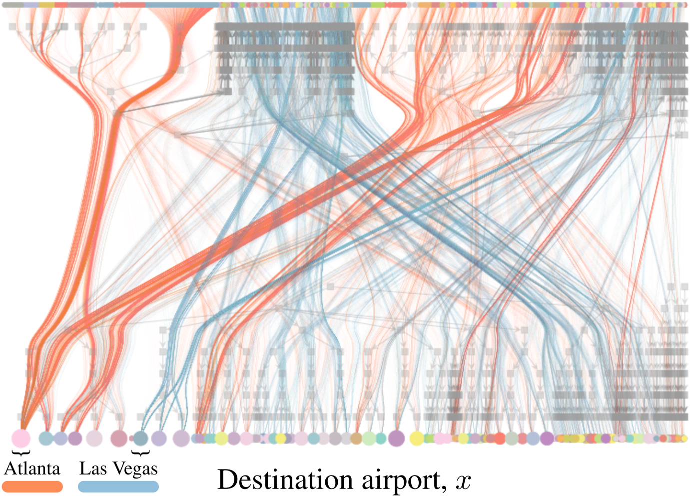
[1] T.P.P. PRL 110 148701 (2013)

n-ORDER MARKOV CHAINS WITH COMMUNITIES

EXAMPLE: FLIGHT ITINERARIES

$$\vec{x}_t = \{x_{t-3}, \text{Altanta} \,|\, \text{Las Vegas}, x_{t-1}\}$$

Previous $n = 3$ airports, $\vec{x}$



Destination airport, x

T. P. P. and Martin Rosvall, arXiv: 1509.04740

Dynamic networks

Each token is an edge: $x_t \to (i,j)_t$

Dynamic network $\to$ Sequence of edges: $\{x_t\} = \{(i,j)_t\}$

Problem: Too many possible tokens! $O(N^2)$

Solution: Group the nodes into B groups.
Pair of node groups $(r,s) \to$ edge group.

Number of tokens: $O(B^2) \ll O(N^2)$

Two-step generative process:

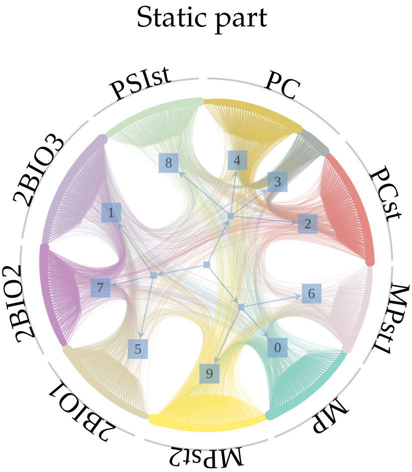
$\{x_t\} = \{(r,s)_t\}$
(n-order Markov chain of pairs of group labels)

$P((i,j)_t|(r,s)_t)$
(static SBM generating edges from group labels)

T. P. P. and Martin Rosvall, arXiv: 1509.04740

Dynamic networks

Example: Student proximity

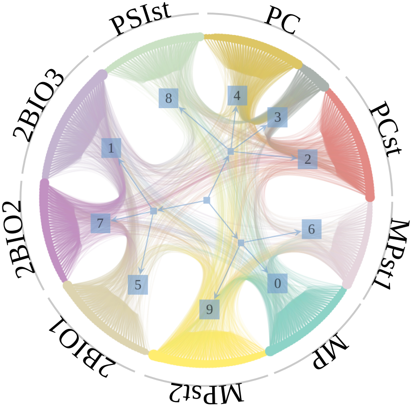


T. P. P. and Martin Rosvall, arXiv: 1509.04740

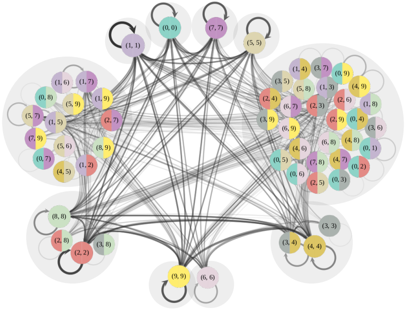
Dynamic networks

Example: Student proximity

Static part



Temporal part



T. P. P. and Martin Rosvall, arXiv: 1509.04740

Dynamic networks in continuous time

$x_\tau \to$ token at continuous time τ

$$P(\{x_\tau\}) = \underbrace{P(\{x_t\})}_{\text{Discrete chain}} \times \underbrace{P(\{\Delta_t\}|\{x_t\})}_{\text{Waiting times}}$$

Exponential waiting time distribution

$$P(\{\Delta_t\}|\{x_t\}, \lambda) = \prod_{\vec{x}} \lambda_{b_{\vec{x}}}^{k_{\vec{x}}} e^{-\lambda_{b_{\vec{x}}} \Delta_{\vec{x}}}$$

Bayesian integrated likelihood

$$\begin{aligned} P(\{\Delta_t\}|\{x_t\}) &= \prod_r \int_0^\infty \mathrm{d}\lambda\ \lambda^{e_r} e^{-\lambda \Delta_r} P(\lambda|\alpha, \beta), \\ &= \prod_r \frac{\Gamma(e_r + \alpha)\beta^\alpha}{\Gamma(\alpha)(\Delta_r + \beta)^{e_r + \alpha}}. \end{aligned}$$

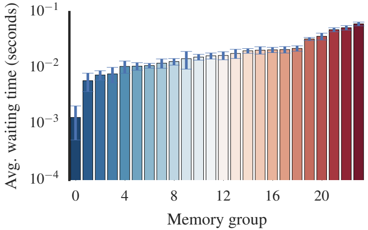
Hyperparameters: α, β. Noninformative limit $\alpha \to 0$, $\beta \to 0$ leads to Jeffreys prior: $P(\lambda) \propto \frac{1}{\lambda}$

T. P. P. and Martin Rosvall, arXiv: 1509.04740

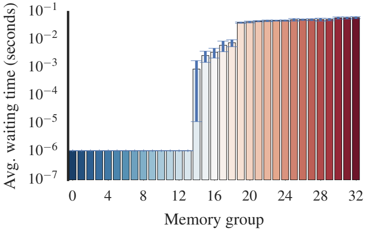
Dynamic networks

Continuous time

$\{x_\tau\} \rightarrow$ Sequence of notes in Beethoven's fifth symphony



Without waiting times
($n = 1$)

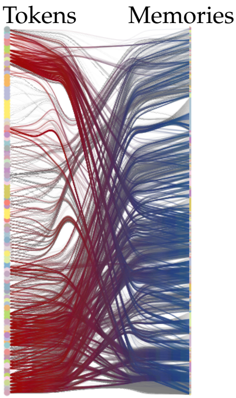


With waiting times
($n = 2$)

NONSTATIONARITY DYNAMIC NETWORKS

$\{x_t\} \rightarrow$ Concatenation of "War and peace," by Leo Tolstoy, and "À la recherche du temps perdu," by Marcel Proust.

Unmodified chain

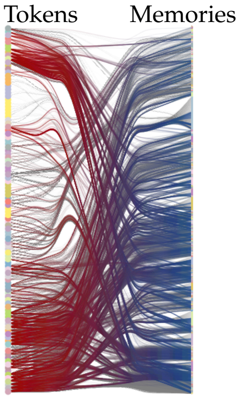


$$-\log_2 P(\{x_t\}, b) = 7,450,322$$

Nonstationarity Dynamic networks

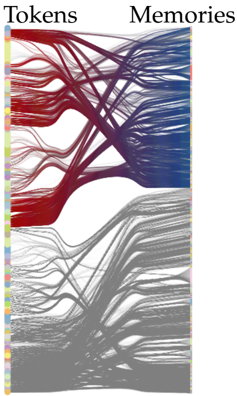
$\{x_t\} \to$ Concatenation of “War and peace,” by Leo Tolstoy, and “À la recherche du temps perdu,” by Marcel Proust.

Unmodified chain



$-\log_2 P(\{x_t\}, b) = 7,450,322$

Annotated chain $x'_t = (x_t, \text{novel})$



$-\log_2 P(\{x_t\}, b) = 7,146,465$

The End

Efficient inference algorithms

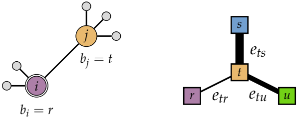
T. P. Peixoto, Phys. Rev. E 89, 012804 (2014)

Smart MCMC

- Choose a random vertex v (happens to belong to block r).
- Move it to a random block $s \in [1, B]$, chosen with a probability $p(r \to s|t)$ proportional to $e_{ts} + \epsilon$, where t is the block membership of a randomly chosen neighbour of v.
- Accept the move with probability

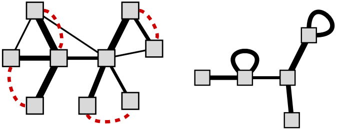
$$a = \min\left\{ e^{-\beta\Delta S} \frac{\sum_t p_t^i p(s \to r|t)}{\sum_t p_t^i p(r \to s|t)}, 1 \right\}.$$

- Repeat.



Fast mixing times.

Agglomerative initialization



Avoids metastable states.

Algorithmic complexity:

$O(N \ln^2 N)$
(independent of B)

Scales up to $10^7 - 10^8$ edges.

graph-tool

Freely available efficient implementation
http://graph-tool.skewed.de