Moving2network : from movement to networks

Vincent Miele

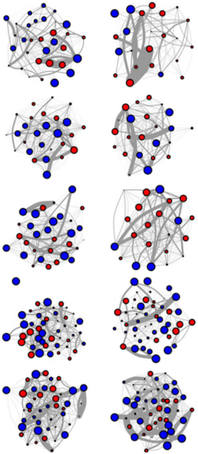
CNRS - Laboratoire Biométrie et Biologie Evolutive

Reinterpreting trajectories

More and more studies using the GPS raw data with a different perspective (network-based).

How can we build networks from GPS tracks ? For what purpose ? How do we study these networks ?

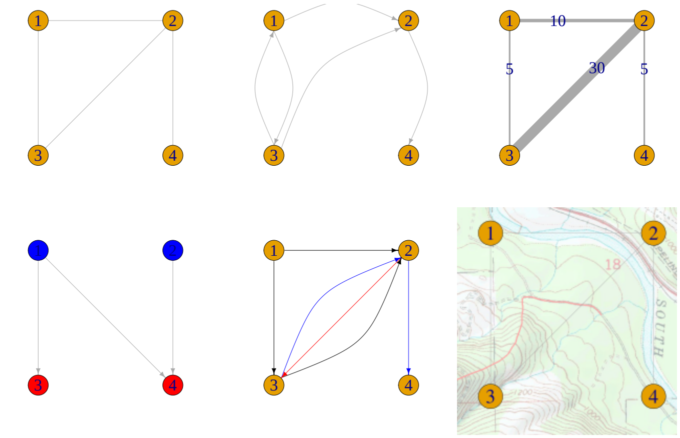
Kays et al , Science (2015)



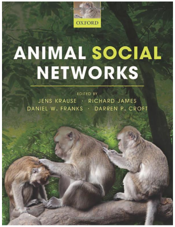
Carter et al , RSOS (2015)

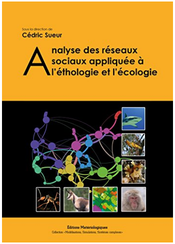
Digression 1 : Networks ?

A *network* is some data represented as a *graph* $\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$ with a set of n *vertices* $\mathbf{V}$ and m *edges* $\mathbf{E}$.



Animal social networks





Sueur et al, Mat. Ed. (2015), Krause et al, Oxford Press (2015)

Animal social networks as a perfect tool to describe the social structure in animal communities.

HOW TO paper by Farine, J. Animal Ecol (2015)

Built from :

1. field observations of association between animals

   e.g. giraffes in Carter *et al*, Anim. Behav. (2013)

2. trapping data

   e.g. field voles in Davis *et al*, J. R. Soc Interface (2015)

3. from sensors-based measurements :
   - proximity loggers

     e.g. badgers with S.Duvillard, LBBE, see Handcock, Sensors (2009)
   - fixed devices distributed in the environment

     e.g. song birds in Farine *et al*, Proc. R. Soc. B (2015), sharks in Jacoby et al, J. R. Soc. Interface (2016)
   - tracking collars (or equivalent)

     e.g. sleepy lizards in Spiegel *et al*, An. Behav. (2015)

Movement to animal social networks

Quoting Noa Pinter-Wollman and colleagues :

Our understanding of animal social networks is largely based on imperfect sampling of associations and interactions. New technologies such as automated tracking systems and proximity data loggers[...] have the potential to close this sampling gap."

Pinter-Wollman et al, Behav Ecol (2014)

Movement to animal social networks

ELSEVIER

Animal Behaviour

Available online 6 December 2017

In Press, Corrected Proof — Note to users

Special Issue: Social Networks

Where should we meet? Mapping social network interactions of sleepy lizards shows sex-dependent social network structure

Orr Spiegel[a], Andrew Sih[a], Stephan T. Leu[b, 1], C. Michael Bull[b]

Show more

https://doi.org/10.1016/j.anbehav.2017.11.001 Get rights and content

Highlights

- We tracked sleepy lizards with GPS loggers to investigate their social network (SN).
- We examined how the observed SN was shaped by social behaviour vs other factors.

- vertices : individuals
- edges : association/contact/proximity inferred from trajectories

Movement to spatial networks

Methods in Ecology and Evolution

Methods in Ecology and Evolution doi: 10.1111/j.2041-210X.2012.00187.x

Developing a deeper understanding of animal movements and spatial dynamics through novel application of network analyses

David M. P. Jacoby[1,2*], Edward J. Brooks[3,4], Darren P. Croft[2] and David W. Sims[1,4]

Table 1. Example hypotheses (*H*) that could be addressed using network analysis of movement data obtained from animal biotelemetry data

H_1	Animals demonstrate repeatable movement patterns/show site fidelity
H_2	Movements and space use differ significantly between time of day/year or between sex/age class
H_3	Environmental variables can be used to predict movement between areas or general movement patterns
H_4	Habitat disturbance at key locations will impact animal movement

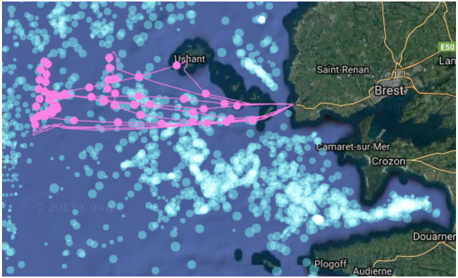
H_5 Delineating spatial units or *network zoning* (Miele, MEE 2014)

Spatial network as a tool to study the movement

- vertices : locations
  habitat patches in Erös, Land. Ecol (2012), turtle ponds in Pereira et al, Lan. Urb. Plan (2011), bat caves with D.Pontier (LBBE), water points with M.Valeix (LBBE), fishway locations in Kirk et al, J.Appl.Eco (2017), acoustic receivers in Jacoby, MEE (2012)
- edges : path (presence/absence) or flows (quantitative) inferred from trajectories

Movement to other networks

Contact networks between fishing boats ?

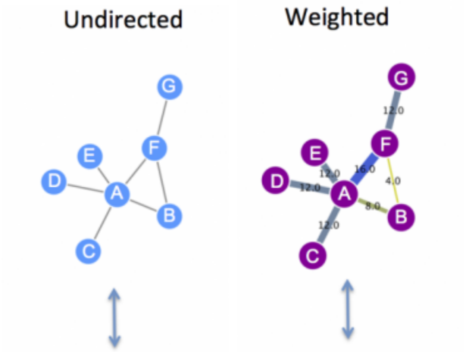


http://globalfishingwatch.org/map

- vertices : boats
- edges : proximity

Digression 2 : adjacency matrix ?

$\{Y_{ij}\}_{1 \leq i,j \leq n}$



$$
\begin{array}{c|ccccccc}
 & A & B & C & D & E & F & G \\
\hline
A & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
B & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
C & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
D & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
E & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
F & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\
G & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}
\qquad
\begin{array}{c|ccccccc}
 & A & B & C & D & E & F & G \\
\hline
A & 0 & 8 & 12 & 12 & 12 & 16 & 12 \\
B & 8 & 0 & 0 & 0 & 0 & 4 & 0 \\
C & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
D & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
E & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\
F & 16 & 4 & 0 & 0 & 0 & 0 & 12 \\
G & 12 & 0 & 0 & 0 & 0 & 12 & 0
\end{array}
$$

(courtesy EMBL)

Building an animal contact network from GPS tracks

Vertices $\equiv$ individuals
Edges $\equiv$?

- $Y_{ij} = (\#$ timestamps $|$ individuals i,j closer than x meters$)/\ \#$ total

Building an animal contact network from GPS tracks

Vertices ≡ individuals
Edges ≡ ?

- Y_{ij} = (# timestamps | individuals i,j closer than x meters)/ # total
- Y_{ij} = (# timestamps | individual j ∈ k-nearest neighbors of i)/ # total

Building an animal contact network from GPS tracks

Vertices ≡ individuals
Edges ≡ ?

- Y_{ij} = (# timestamps | individuals i,j closer than x meters)/ # total
- Y_{ij} = (# timestamps | individual j $\in$ k-nearest neighbors of i)/ # total
- $Y'_{ij} = \mathbb{1}_{Y_{ij} > \epsilon}$ where ϵ is some threshold (quantile of Y distribution, mixture, percolation threshold,...)

Demo (part 1)

Trajectories of 25 baboons.

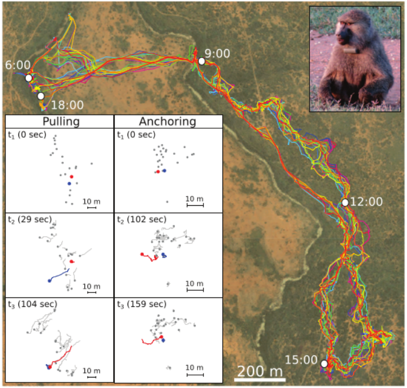


Fig. 1. Extracting pulls and anchors from movement data. Baboon trajectories (25 individuals)

Strandburg-Peskin et al, Science (2015) ; GPS data (1 position/sec) available at

`https://www.datarepository.movebank.org/handle/10255/move.405`

R packages `igraph` and `dynsbm`

Network measures

Quoting Mark Newman :

Much of the current research on networks [...] aimed at answering the question "How can we tell what a network looks like, when we can't actually look at it ?"

Newman and Leicht, PNAS (2008)

→ descriptive statistics ≡ *network measures*

Networks measures

Many network measures well defined in the binary case :

Newman, Oxford Press (2010)

- *density* or *connectance* : $\frac{2m}{n(n-1)}$
- *degree* centrality of vertex i : $d_i = \sum_k Y_{ik}$
- degree distribution (hubs)
- eigenvector centrality of vertex i : $c_i = \sum_j Y_{ik} c_k$

Networks measures

Many network measures well defined in the binary case :

Newman, Oxford Press (2010)

- *density* or *connectance* : $\frac{2m}{n(n-1)}$
- *degree* centrality of vertex i : $d_i = \sum_k Y_{ik}$
- degree distribution (hubs)
- eigenvector centrality of vertex i : $c_i = \sum_j Y_{ik} c_k$

$+$ other measures based on *paths*
(*e.g. betweenness* and *closeness* centrality ; only if studying spreading processes)

Networks measures

Many network measures well defined in the binary case :

Newman, Oxford Press (2010)

- *density* or *connectance* : $\frac{2m}{n(n-1)}$
- *degree* centrality of vertex i : $d_i = \sum_k Y_{ik}$
- degree distribution (hubs)
- eigenvector centrality of vertex i : $c_i = \sum_j Y_{ik} c_k$

$+$ other measures based on *paths*
(*e.g.* *betweenness* and *closeness* centrality ; only if studying spreading processes)

⚠ network measures extended to the weighted cases, but take caution

Beyond networks measures

Animal societies are known to be :

- modular (animals gathered in small groups)
- highly structured with individuals playing different roles (hierachy and leadership, ...)

→ finding *modules* or *communities* ≡ most edges (binary case) or a high proportion of the edge weight (weighted case) is within rather than between groups

Fortunato, Physics reports (2010,2016)

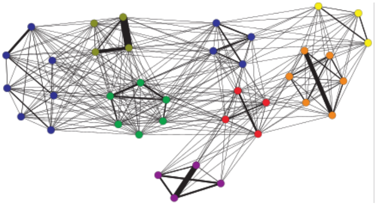
Beyond networks measures

Animal societies are known to be :

- modular (animals gathered in small groups)
- highly structured with individuals playing different roles (hierachy and leadership, ...)

→ finding *modules* or *communities* ≡ most edges (binary case) or a high proportion of the edge weight (weighted case) is within rather than between groups

Fortunato, Physics reports (2010,2016)

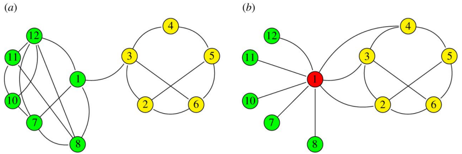


Verreaux's sifakas in Springer et al, J. Anim. Eco. (2017)

Beyond networks measures

Animal societies are known to be :

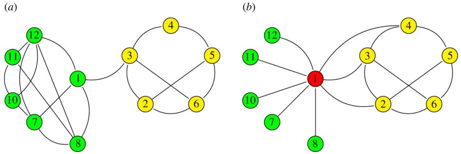
- modular (animals gathered in small groups)
- highly structured with individuals playing different roles (hierachy and leadership, ...)



Beyond networks measures

Animal societies are known to be :

- modular (animals gathered in small groups)
- highly structured with individuals playing different roles (hierachy and leadership, ...)



→ finding groups of vertices that have similar interaction patterns (hubs, modules, peripherical vertices,...) without *a priori* (*i.e. statistical inference*)

Stochastic block models (SBM)

SBM :

1. grouping vertices that have similar connectivity patterns
2. modelling explicitely the weight distribution
3. more compliant with FP/FN edges (probabilistic framework)

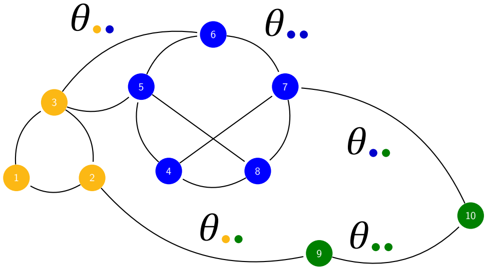
Newman and Leicht , PNAS (2008) Daudin, Picard, Robin, Stat. Comp (2008) Guimera and Sales-Pardo, PNAS (2009)

Zanghi et al, Pattern Recognition (2008), Ann. Appl. Stat. (2010) Picard et al, BMC Bioinformatics (2009) Kéfi et al, PLoS Biol. (2016)

SBM - notations

- Observe Y adjacency matrix
- Presence/absence of an edge $Y = (Y_{ij})_{1 \leq i,j \leq n}$
  binary, discrete or continuous values

SBM



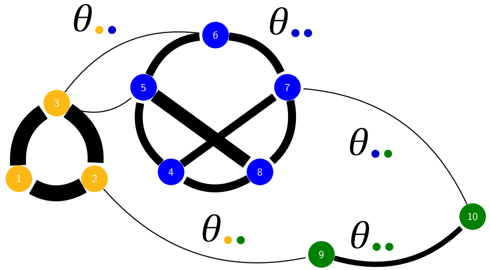
$n = 10, Q = 3,$

$Z_5 = \bullet,$

$Y_{12} = 1, Y_{15} = 0$

- Q groups (=colors $\bullet\bullet\bullet$) in proportion $\boldsymbol{\alpha} = \{\alpha_q\}_{1 \le q \le Q}$
- $\{Z_i\}_{1 \le i \le n}$ i.i.d. in $\{1, \ldots, Q\}$, group membership not observed
- the r.v. $Y_{ij} \mid Z_i = \bullet, Z_j = \bullet \sim f(\cdot\,; \theta_{\bullet\bullet})$

SBM



$n = 10, Q = 3,$

$Z_5 = \bullet,$

Y_{12}, Y_{15} binary, discrete or continuous

- the r.v. $Y_{ij} \mid Z_i = \bullet, Z_j = \bullet \sim f(\cdot\, ; \theta_{\bullet\bullet})$

  f is any probability distribution parametrised by θ :
  - Bernouilli distribution in the binary case
  - non-parametric distribution in the discrete case
  - 0-inflated Gaussian distribution in the continuous case

SBM : estimation

1. model estimation performed with a EM-like optimization algorithm

$$\log \mathbb{P}_\theta(\mathbf{Y},\mathbf{Z}) = \sum_i \sum_q \mathbb{1}_{Z_i=q} \log \alpha_q + \sum_{i,j} \sum_{q,l} \mathbb{1}_{Z_i=q} \mathbb{1}_{Z_j=l} \log f(Y_{ij};\theta_{ql}).$$

2. model selection framework to choose the number of groups

Demo (part 2)

Trajectories of 25 baboons.

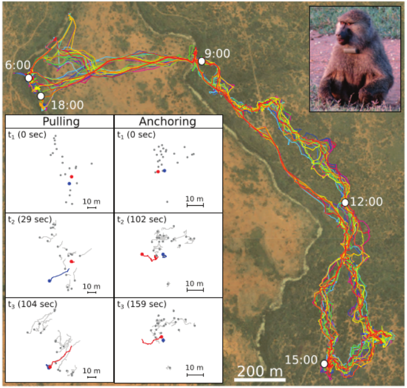


Fig. 1. Extracting pulls and anchors from movement data. Baboon trajectories (25 individuals)

Strandburg-Peskin et al, Science (2015) ; GPS data (1 position/sec) available at

`https://www.datarepository.movebank.org/handle/10255/move.405`

R packages `igraph` and `dynsbm`

Interpretation

How do we interpret the clustering results ? Behavior or ecological factors ? Caution is required...

Quoting Noa Pinter-Wolman :

For behavioral ecologists, many standard [clustering] algorithms still provide an incomplete understanding of spatial drivers because they do not use spatiotemporal data per se [and can not] provide a more complete understanding of how putatively "social" networks depend on, or can be distinguished from, these underlying ecological factors.

Extending an animal contact network from different data

Studying the interplay between individuals of different species, using GPS tracks :

- colored vertices
- edges in a *multilayer* network (intra+interspecies)

Extending an animal contact network from different data

Studying the interplay between individuals of different species, using GPS tracks :

- colored vertices
- edges in a *multilayer* network (intra+interspecies)

Studying different kind of interaction, using loggers (accelerometers, sound-recording devices,...)

- colored edges in a *multiplex* network

  Courbin et al, Eco. Mono. (2014)

Extending an animal contact network from different data

Studying the interplay between individuals of different species, using GPS tracks :

- colored vertices
- edges in a *multilayer* network (intra+interspecies)

Studying different kind of interaction, using loggers (accelerometers, sound-recording devices,...)

- colored edges in a *multiplex* network

  Courbin et al, Eco. Mono. (2014)

Studying multiple independent networks at different locations

- comparing network structure

  in a beta-diversity framework, Ohlmann et al, in prep

Building a series of animal contact networks from GPS tracks

Behavioral Ecology
The official journal of the ISBE
International Society for Behavioral Ecology
doi:10.1093/beheco/arx047

Invited Review

The dynamics of animal social networks: analytical, conceptual, and theoretical advances

Noa Pinter-Wollman,[a] Elizabeth A. Hobson,[b] Jennifer E. Smith,[c] Andrew J. Edelman,[d] Daizaburo Shizuka,[e,f] Shermin de Silva,[g] James S. Waters,[h] Steven D. Prager,[i] Takao Sasaki,[h] George Wittemyer,[g] Jennifer Fewell,[h] and David B. McDonald[i]

Received: 6 June 2017 | Accepted: 1 October 2017
DOI: 10.1111/1365-2656.12764
Journal of Animal Ecology

'HOW TO...' PAPER

When to choose dynamic vs. static social network analysis

Damien R. Farine[1,2,3]

Why do we aggregate all the information whereas the GPS tracks are time-ordered ?

If there is a social structure (at least a non-random organization), what is its dynamics ?

(*e.g.* fission-fusion societies in zebra and onagers ; Rubenstein, PloS ONE (2015))

How does it vary with other factors ? (landscape changes, response to stress)

Is it stable/variable with individual turnover ?

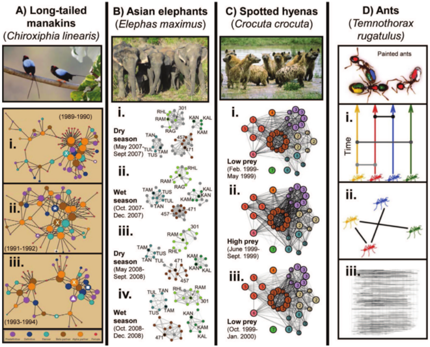
Building a dynamic animal contact network from GPS tracks

An approach to study the temporal dynamics of contact networks is the discrete *snapshots* approach :

1. data over time are aggregated within relevant intervals (expert-based)
2. one snapshot of the contact network for each interval

→ *i.e.* a *dynamic* network.

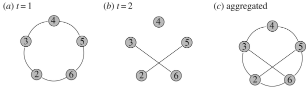
Digression 3 : Static vers dynamic versus temporal



Pinter-Wollman, Behav. Ecol. (2013)

Viewing same data as :

- a *static* or *aggregated* network
- a *dynamic* network



- a *temporal* (or *time-ordered*) network

  Holme, Europ. Phys. J. B (2015)

Networks measures and community structures

Networks measures : extensions to dynamic network or time series of measures

Community structure in dynamic networks : still in progress

Mucha et al, Science (2010) ; review in Cazabet, arxiv (2017))

What do we want to model in a dynamic animal contact network ?

Mission statement for an extension of SBM :

- recontruct the whole story without label switching
- allow for group rearrangement / group switches
- accept group (dis)appearance
- cope with presence/absence of individuals

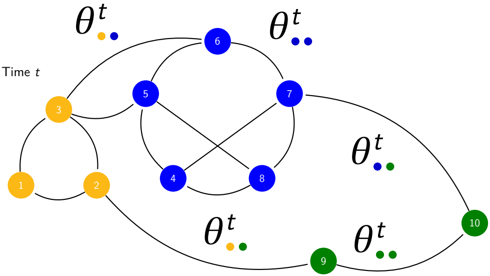
Matias and Miele, J. Royal. Soc. Series B (2017)

Miele and Matias, RSOS (2017)

dynSBM - notations

- Observe $Y^1, \ldots, Y^T$ adjacency matrices (the "snapshots")
- Presence/absence of an edge $\forall t, Y^t = (Y^t_{ij})_{1 \leq i,j \leq n_t}$
  binary, discrete or continuous values
- Individuals may be present/absent at each time step t

dynSBM - a SBM for each time step



$n_t = 10, Q = 3,$

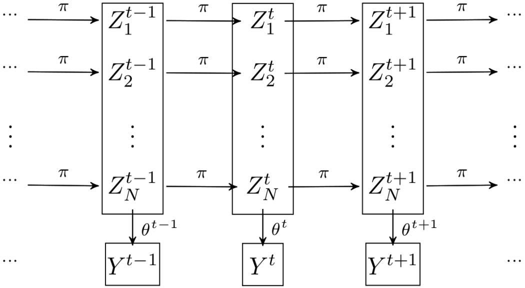
$Z_5^t = \bullet,$

$Y_{12}^t = 1, Y_{15}^t = 0$

- Q groups (=colors $\bullet\bullet\bullet$).
- $\{Z_i^t\}_{1 \leq i \leq n}$ i.i.d. in $\{1, \ldots, Q\}$, group membership not observed
- the r.v. $Y_{ij}^t \mid Z_i^t = \bullet, Z_j^t = \bullet \sim f(\cdot\ ; \theta^t_{\bullet\bullet})$

dynSBM - markov chains for the group dynamics

- Across time : each $Z_i = (Z_i^t)_{1 \leq t \leq T}$ is a stationary Markov chain on $\{1, \ldots, Q\}$ with transition $\boldsymbol{\pi} = (\pi_{qq'})_{1 \leq q, q' \leq Q}$
- Across individuals : $(Z_i)_{1 \leq i \leq N}$ iid



Identifiability : the problem

If both $(\theta^t)_t$ and $(Z^t)_t$ can change, the parameters are not identifiable because of the **label switching** phenomenon.

Identifiability : the problem

If both $(\theta^t)_t$ and $(Z^t)_t$ can change, the parameters are not identifiable because of the **label switching** phenomenon.

Example : 2 identical snapshots with 2 groups.
$t = 1$, group 1 $\equiv$ hubs and group 2 $\equiv$ periphery
$t = 2$, group 2 $\equiv$ hubs and group 1 $\equiv$ periphery
or group 1 $\equiv$ hubs and group 2 $\equiv$ periphery ?

Same likelihood, different parameters, different interpretation...

Identifiability : the problem

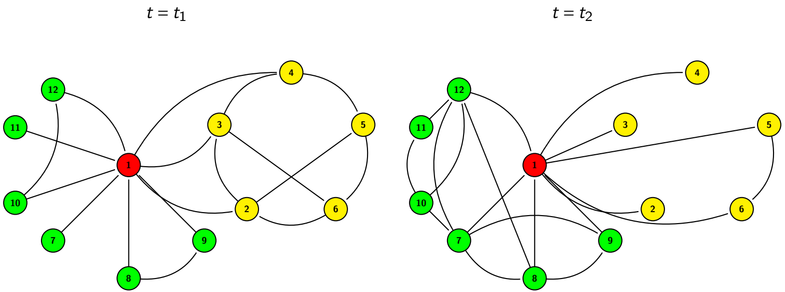
If both $(\theta^t)_t$ and $(Z^t)_t$ can change, the parameters are not identifiable because of the **label switching** phenomenon.

Example : 2 identical snapshots with 2 groups.
$t = 1$, group 1 $\equiv$ hubs and group 2 $\equiv$ periphery
$t = 2$, group 2 $\equiv$ hubs and group 1 $\equiv$ periphery
or group 1 $\equiv$ hubs and group 2 $\equiv$ periphery ?

Same likelihood, different parameters, different interpretation...

$\rightarrow$ we must add constraints.

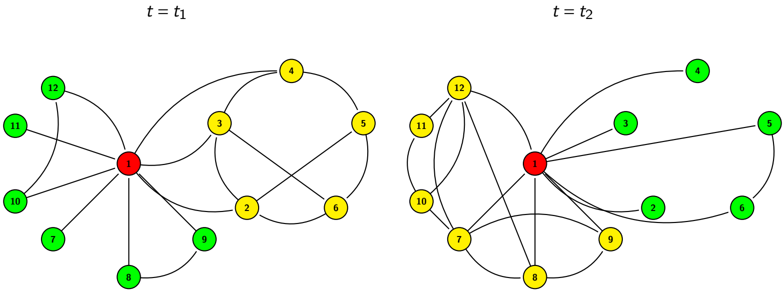
Identifiability : the solution



First scenario : constraints on the groups memberships (stable in time)

- $\forall i, Z_i^1 = Z_i^2$
- For $t = t_1$, $\theta^{t_1}_{\bullet,\bullet} =$ periphery and $\theta^{t_1}_{\bullet,\bullet} =$ community ,
- For $t = t_2$, $\theta^{t_2}_{\bullet,\bullet} =$ community and $\theta^{t_2}_{\bullet,\bullet} =$ periphery.

Identifiability : the solution



Second scenario : constraints on the groups parameters (stable in time)

- $\forall 2 \leq i \leq 6, Z_i^1 =$ community, $Z_i^2 =$ periphery
- $\forall 7 \leq i \leq 12, Z_i^1 =$ periphery, $Z_i^2 =$ community
- $(\theta^{t_1}, \gamma^{t_1}) = (\theta^{t_2}, \gamma^{t_2})$

dynSBM : our proposal

1. combines SBMs and Markov chains
2. identifiability is guaranteed by $\forall q, \forall t, t'$, $\theta^t_{qq} = \theta^{t'}_{qq}$
3. model estimation performed with a EM-like optimization algorithm

$$\log \mathbb{P}_\theta(\mathbf{Y}, \mathbf{Z}) = \sum_i \sum_q \mathbb{1}_{Z_i^1 = q} \log \alpha_q + \overbrace{\sum_{t=2}^{T} \sum_i \sum_{q,q'} \mathbb{1}_{Z_i^{t-1}=q} \mathbb{1}_{Z_i^t = q'} \log \pi_{qq'}}^{markov\ \ chain}$$
$$+ \underbrace{\sum_{t=1}^{T} \sum_{i,j} \sum_{q,l} \mathbb{1}_{Z_i^t = q} \mathbb{1}_{Z_j^t = l} \log f(Y_{ij}^t ; \theta_{ql}^t)}_{SBM\ \ \forall t}.$$

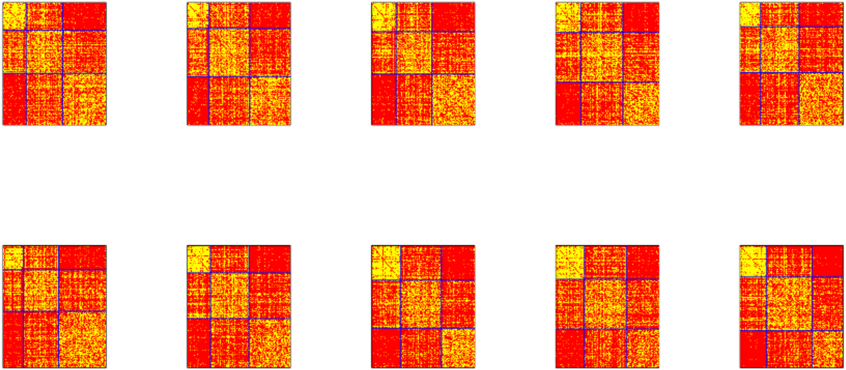
4. model selection framework to choose the number of groups (in progress)

Matias and Miele, J. Royal. Soc. Series B (2017)

Miele and Matias, RSOS (2017)

dynSBM : output

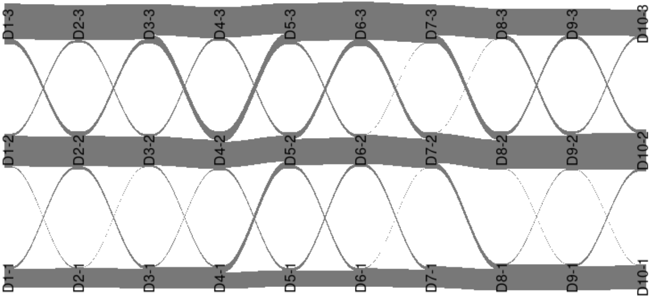
152 ants followed during 10 days
3 groups found



Mersch et al, Science (2013)

dynSBM : output

152 ants followed during 10 days



3 social groups : *nurses, foragers* and *cleaners*

Mersch et al, Science (2013), Miele and Matias, RSOS (2017)

Demo (part 3)

Trajectories of 25 baboons.

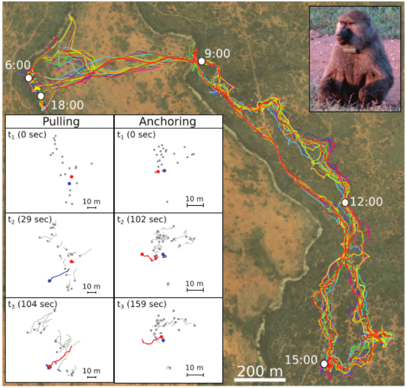


Fig. 1. Extracting pulls and anchors from movement data. Baboon trajectories (25 individuals)

Strandburg-Peskin et al, Science (2015) ; GPS data (1 position/sec) available at

`https://www.datarepository.movebank.org/handle/10255/move.405`

R packages `igraph` and `dynsbm`

Interpretation

How do we interpret the clustering results ?

Superimpose different times series, including the time series with the groups evolving in time

My contact network

Christophe Bonenfant (LBBE)

Catherine Matias (LPMA)

Stéphane Dray (LBBE)

Franck Picard (LBBE)

Marion Valeix (LBBE)

Stéphane Robin (Agroparistech)