From Phylogenetic to Bayesian Networks

Benjamin Teo[1], Paul Bastide[2], Cécile Ané[1,3]

[1] Department of Statistics, University of Wisconsin-Madison
[2] IMAG, Université de Montpellier, CNRS
[3] Department of Botany, University of Wisconsin-Madison

Statistiques au Sommet de Rochebrune, Mars 2024

Polemonium

(Teo et al., 2023)

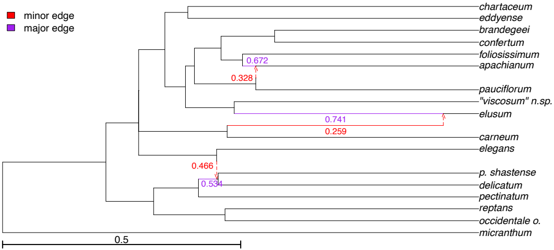


Figure: *Teo et al. (2023)*

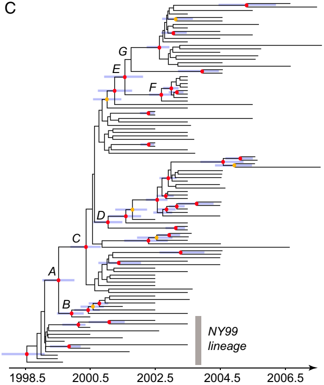
Polemonium eddyense

Polemonium reptens

Polemonium carneum

West Nile Virus (WNV)

(Pybus et al., 2012)



Coronaviruses

(Müller et al., 2022)

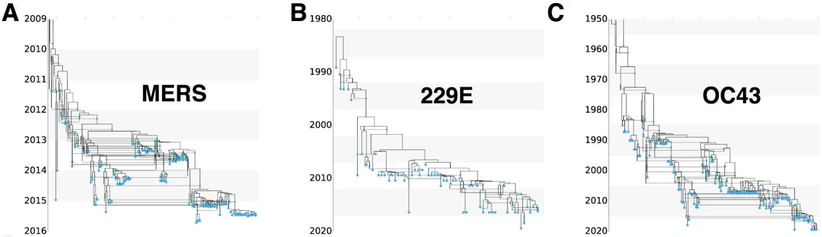


Figure: Müller et al. (2022)

Native American and Arctic populations (Nielsen et al., 2023)

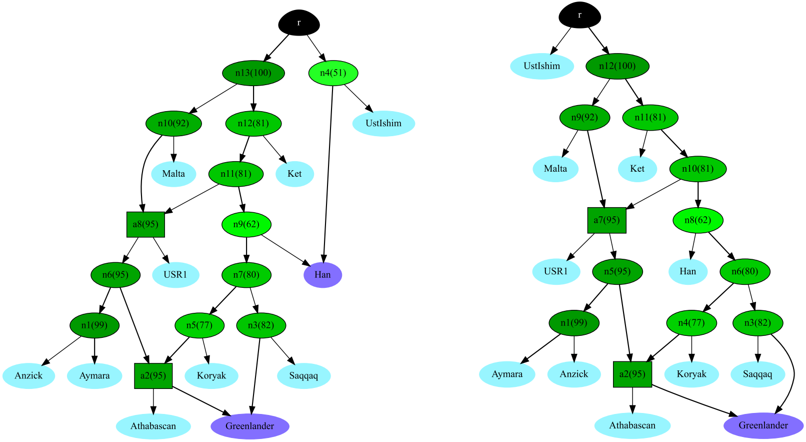
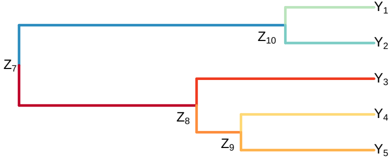
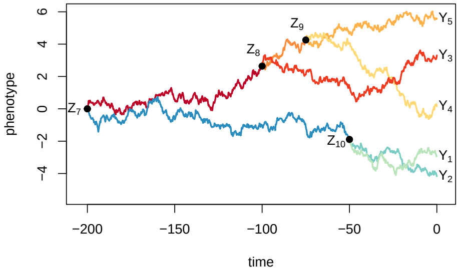


Figure: Nielsen et al. (2023)

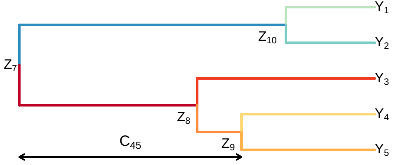
Brownian Motion on a Tree

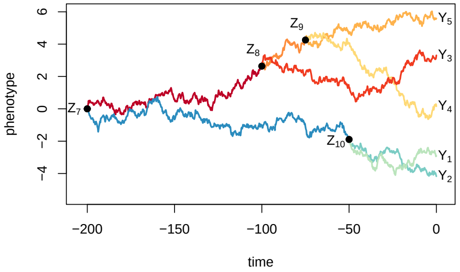
(Felsenstein, 1985)





Brownian Motion on a Tree (Felsenstein, 1985)

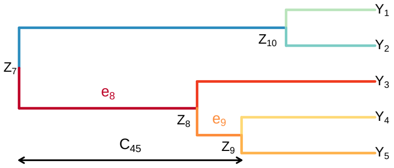


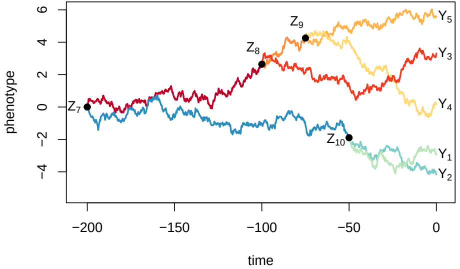


BM Variance: Shared Evolution Time $\mathbb{C}\text{ov}[Y_4; Y_5] = \sigma^2 C_{45}$

Brownian Motion on a Tree

(Felsenstein, 1985)



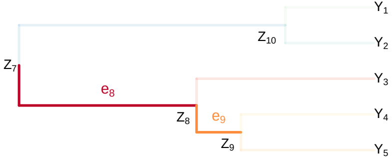


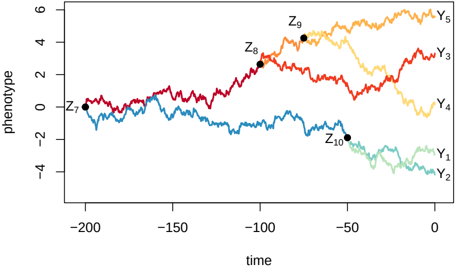
BM Variance: Shared Evolution Time

$$\mathbb{C}\mathrm{ov}[Y_4; Y_5] = \sigma^2 C_{45}$$

$$C_{45} = \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e$$

Brownian Motion on a Tree (Felsenstein, 1985)





BM Variance: Shared Evolution Time $\mathbb{C}\text{ov}[Y_4; Y_5] = \sigma^2 C_{45}$

$$C_{45} = \ell_{e_8} + \ell_{e_9} = \sum_{e \in \{e_8, e_9\}} \ell_e = \sum_{e \in p_4 \cap p_5} \ell_e$$

p_i: path from root to tip i:

$$p_4 = \{e_8, e_9, e_4\}$$
$$p_5 = \{e_8, e_9, e_5\}$$

Hybrids

Hybrids

Leopard

Hybrids

Leopard

+

Lion

Hybrids

Leopard

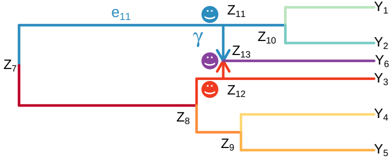
$+$

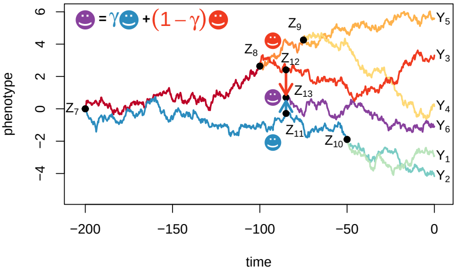
Lion

$=$

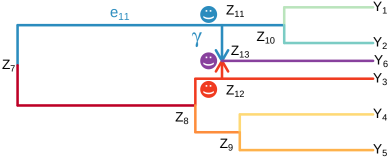
Leopon

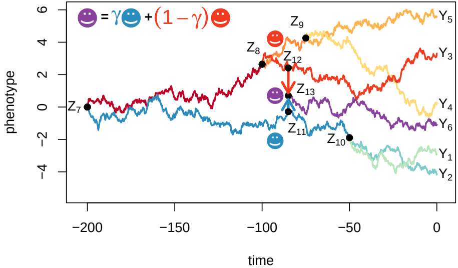
Brownian Motion on a Network





Brownian Motion on a Network



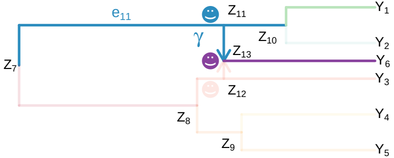


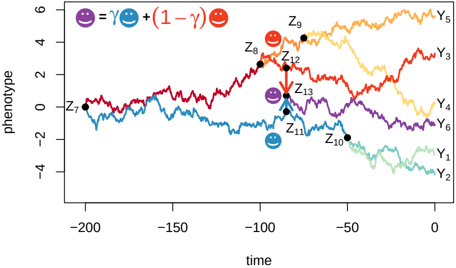
BM Variance: "Shared Evolution Time"

Problem Path to Y_6 not unique

$$C_{ij}^{\text{tree}} = \sum_{e \in p_i \cap p_j} \ell_e$$

Brownian Motion on a Network



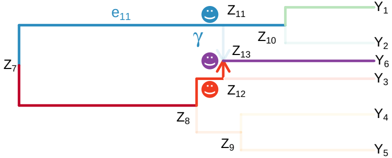


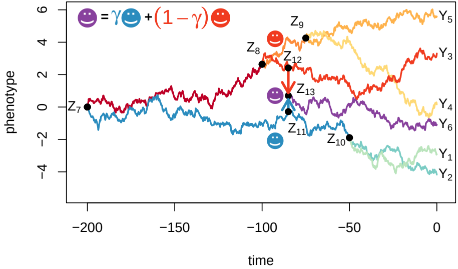
BM Variance: “Shared Evolution Time”

$$C_{16} = \ell_{e_{11}} \quad \times \gamma$$

$$\mathcal{P}_6 = \left\{ \begin{array}{l} \{e_{11}, e_{13}, e_6\}, \\ \\ \end{array} \right\}$$

Brownian Motion on a Network



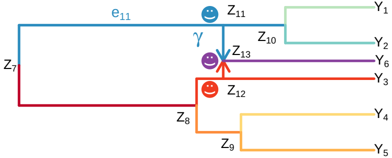


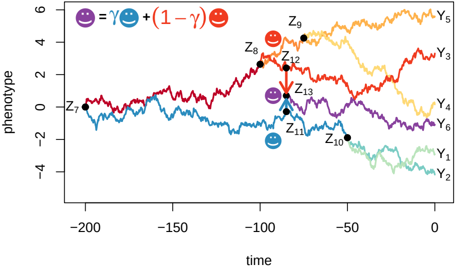
BM Variance: "Shared Evolution Time"

$$C_{16} = \ell_{e_{11}} \times \gamma \\ + 0 \times (1-\gamma)$$

$$\mathcal{P}_6 = \left\{ \begin{array}{l} \{e_{11}, e_{13}, e_6\}, \\ \{e_8, e_{12}, e_{12}, e_6\} \end{array} \right\}$$

Brownian Motion on a Network





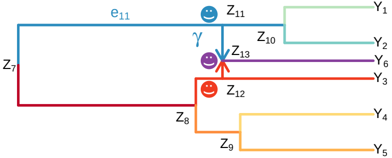
BM Variance: “Shared Evolution Time”

$$C_{16} = \ell_{e_{11}} \times \gamma + 0 \times (1-\gamma)$$

$$C_{ij}^{\text{net}} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \Big(\prod_{e \in p_i} \gamma_e\Big)\Big(\prod_{e \in p_j} \gamma_e\Big) \sum_{e \in p_i \cap p_j} \ell_e$$

Pre-Order

Distribution



$$C_{ij}^{net} = \sum_{\substack{p_i \in \mathcal{P}_i \\ p_j \in \mathcal{P}_j}} \Big(\prod_{e \in p_i} \gamma_e\Big)\Big(\prod_{e \in p_j} \gamma_e\Big) \sum_{e \in p_i \cap p_j} \ell_e$$

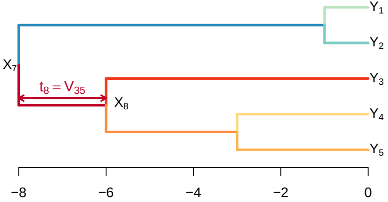
BM on a Network

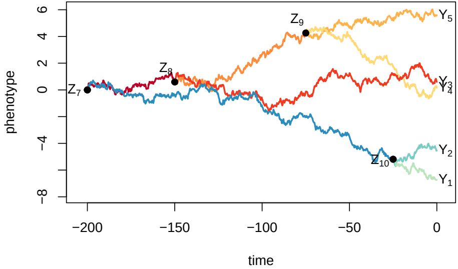
$$\mathbf{Y} = \mu \mathbf{1} + \sigma \mathbf{E} \qquad \mathbf{E} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{Y,Y})$$

Inference: $\hat{\mu}$, $\hat{\sigma}$
Ancestral State Reconstruction: $p(Z|Y)$

Formulas: need to invert (non sparse) matrix $\mathbf{C}_{Y,Y}$

Heredity rules

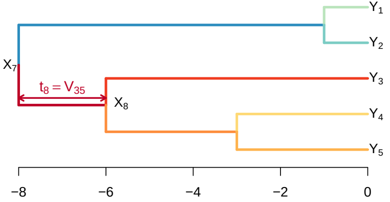


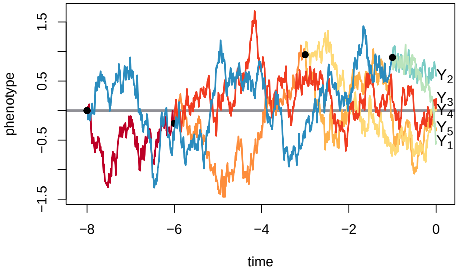


Brownian Motion (Felsenstein, 1985)

$$X_i \mid X_{\text{pa}(i)} \sim \mathcal{N}(X_{\text{pa}(i)}; \sigma^2 \ell_i)$$

Heredity rules

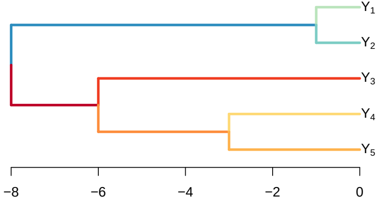


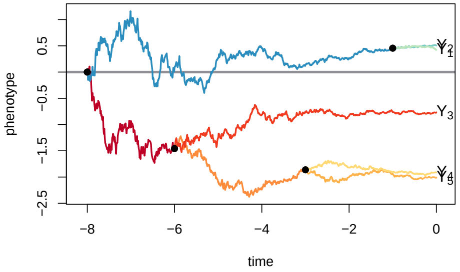


Ornstein-Uhlenbeck (Hansen, 1997)

$$X_i \mid X_{\mathrm{pa}(i)} \sim \mathcal{N}(e^{-\alpha\ell_i} X_{\mathrm{pa}(i)} + (1 - e^{-\alpha\ell_i})\beta_i; \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\ell_i}))$$

Heredity rules

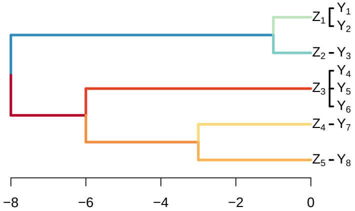


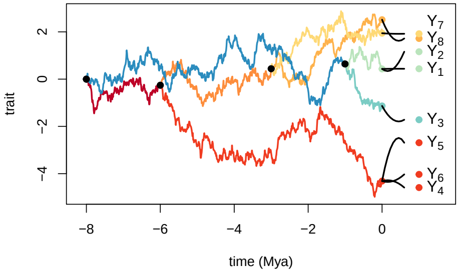


Early Burst (Harmon et al., 2010)

$$X_i \mid X_{\text{pa}(i)} \sim \mathcal{N}(X_{\text{pa}(i)}; \frac{\sigma^2}{r} e^{rt_{\text{pa}(i)}}(e^{r\ell_i} - 1))$$

Heredity rules

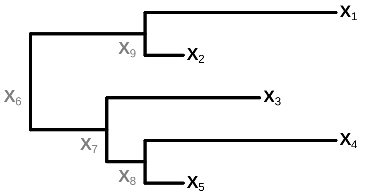




Within species variance

$$X_i \mid X_{\mathrm{pa}(i)} \sim \mathcal{N}(X_{\mathrm{pa}(i)}; \sigma_e^2)$$

General Model on a Tree



$$\mathbf{X}_r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

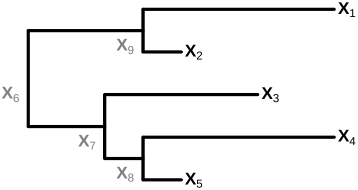
$$\mathbf{X}_j \mid \mathbf{X}_{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{X}_{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\beta_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Measurement errors, drift, shifts, Integrated OU...

General Model on a Tree



$$\mathbf{X}_r \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma}) \quad \text{root}$$

$$\mathbf{X}_j \mid \mathbf{X}_{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{X}_{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right) \quad \text{nodes}$$

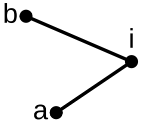
BM: $\mathbf{q}_j = \mathbf{I}_p$, $\mathbf{r}_j = \mathbf{0}_p$, $\boldsymbol{\Sigma}_j = \ell_j \mathbf{R}$.

OU: $\mathbf{q}_j = e^{-\mathbf{A}\ell_j}$, $\mathbf{r}_j = (\mathbf{I}_p - e^{-\mathbf{A}\ell_j})\beta_j$, $\boldsymbol{\Sigma}_j = \mathbf{S} - e^{-\mathbf{A}\ell_j}\mathbf{S}e^{-\mathbf{A}^T\ell_j}$.

Measurement errors, drift, shifts, Integrated OU...

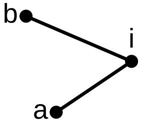
Question: what happens at hybrids ?

Hybridizations



$$\mathbf{X}_i = \gamma_a(\mathbf{q}_a\mathbf{X}_a + \mathbf{r}_a + \boldsymbol{\epsilon}_a) \qquad \boldsymbol{\epsilon}_a \sim \mathcal{N}(0, \boldsymbol{\Sigma}_a)$$
$$+ \gamma_b(\mathbf{q}_b\mathbf{X}_b + \mathbf{r}_b + \boldsymbol{\epsilon}_b) \qquad \boldsymbol{\epsilon}_b \sim \mathcal{N}(0, \boldsymbol{\Sigma}_b)$$

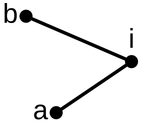
Hybridizations



$$\mathbf{X}_i = \gamma_a(\mathbf{q}_a\mathbf{X}_a + \mathbf{r}_a + \boldsymbol{\epsilon}_a) \qquad \boldsymbol{\epsilon}_a \sim \mathcal{N}(0, \boldsymbol{\Sigma}_a)$$
$$+ \gamma_b(\mathbf{q}_b\mathbf{X}_b + \mathbf{r}_b + \boldsymbol{\epsilon}_b) \qquad \boldsymbol{\epsilon}_b \sim \mathcal{N}(0, \boldsymbol{\Sigma}_b)$$

$$\mathbf{X}_i = \begin{pmatrix} \gamma_a\mathbf{q}_a & \mathbf{0} \\ \mathbf{0} & \gamma_b\mathbf{q}_b \end{pmatrix} \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix} \quad + \quad \gamma_a\mathbf{r}_a + \gamma_b\mathbf{r}_b \quad + \quad \gamma_a\boldsymbol{\epsilon}_a + \gamma_b\boldsymbol{\epsilon}_b$$

Hybridizations



$$\mathbf{X}_i = \gamma_a(\mathbf{q}_a\mathbf{X}_a + \mathbf{r}_a + \boldsymbol{\epsilon}_a) \qquad \boldsymbol{\epsilon}_a \sim \mathcal{N}(0, \boldsymbol{\Sigma}_a)$$
$$+ \gamma_b(\mathbf{q}_b\mathbf{X}_b + \mathbf{r}_b + \boldsymbol{\epsilon}_b) \qquad \boldsymbol{\epsilon}_b \sim \mathcal{N}(0, \boldsymbol{\Sigma}_b)$$

$$\mathbf{X}_i = \begin{pmatrix} \gamma_a\mathbf{q}_a & \mathbf{0} \\ \mathbf{0} & \gamma_b\mathbf{q}_b \end{pmatrix} \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix} \quad + \quad \gamma_a\mathbf{r}_a + \gamma_b\mathbf{r}_b \quad + \quad \gamma_a\boldsymbol{\epsilon}_a + \gamma_b\boldsymbol{\epsilon}_b$$

$$\mathbf{X}_j \mid \mathbf{X}_{\mathrm{pa}(j)} \sim \mathcal{N}\left(\mathbf{q}_j\mathbf{X}_{\mathrm{pa}(j)} + \mathbf{r}_j,\ \boldsymbol{\Sigma}_j\right)$$

Graphical Model

DAG: Trait evolution → linear Gaussian distribution

Factors: node v givent its parents u

$$\phi_v(X_v|X_u, \theta; u \in \text{pa}(v)) \sim \mathcal{N}\left(\mathbf{q}_j \mathbf{X}_{\text{pa}(j)} + \mathbf{r}_j,\ \mathbf{\Sigma}_j\right)$$

Joint Distribution:

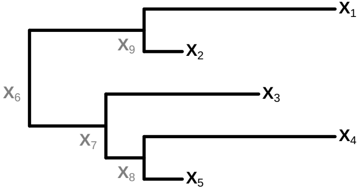
$$p_\theta(X_v; v \in V) = \prod_{v \in V} \phi_v(X_v|X_u, \theta; u \in \text{pa}(v))$$

Goal: compute

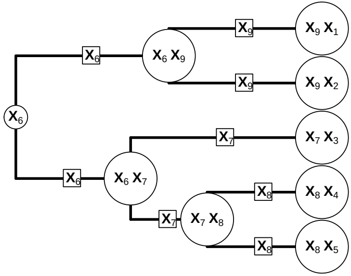
- likelihood $p_\theta(X_o; o \in V_{obs})$
- ancestral reconstruction $p_\theta(X_v|X_o; o \in V_{obs})$

Tool: Belief Propagation

Clique Tree



Phylogenetic Tree



Clique Tree

Cluster graph:

- Nodes: clusters C_i (family preserving)
- Edges: Sepsets $S_{ij} \subseteq C_i \cap C_j$
- Running intersection: each variable defines a tree

Belief Propagation (Koller and Friedman, 2009)

Initialization:

- $\beta_i = \prod_{v:\text{scope}(v)\in C_i} \phi_v$ — cluster beliefs
- $\mu_{i,j} = 1$ — sepset beliefs

Message passing: from C_i to C_j

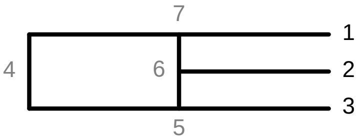
- $\tilde{\mu}_{i,j} = \int_{C_i \setminus S_{i,j}} \beta_i d(C_i \setminus S_{i,j})$
- $\beta_j \leftarrow \beta_j \frac{\tilde{\mu}_{i,j}}{\mu_{i,j}}$
- $\mu_{i,j} \leftarrow \tilde{\mu}_{i,j}$

At calibration:

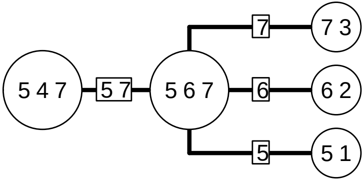
- $\beta_i = p(X_v, v \in C_i | X_o, o \in V_{obs})$
- $\mu_{i,j} = p(X_v, v \in S_{i,j} | X_o, o \in V_{obs})$

Clique tree: only two traversals are needed.

Clique Tree



Phylogenetic Network



Clique Tree

Clique tree construction:

- Moralization, triangulation, maximum spanning tree

BP:

- Complexity depends on max clique size

Implementation

Linear Gaussian

- Canonical form: $\mathrm{C}(\mathbf{x}; \mathbf{K}, \mathbf{h}, g) = \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{K}\mathbf{x} + \mathbf{h}^T\mathbf{x} + g\right)$
- Message passing: simple matrix operations

julia package `PhyloGaussianBeliefProp.jl`

- Interface for phylogenetic neworks
- Clique tree construction
- Likelihood and Ancestral state computation with BP
- Optimization using `ForwardDiff.jl`

Gradient Computation

Fisher's Identity — Cappé et al. (2005)

$$\nabla_{\theta'} [\log p_{\theta'}(\mathbf{Y})]|_{\theta'=\theta} = \mathbb{E}_\theta \left[\nabla_{\theta'} [\log p_{\theta'}(\mathbf{X}_v; v \in V)]|_{\theta'=\theta} \mid \mathbf{Y} \right]$$

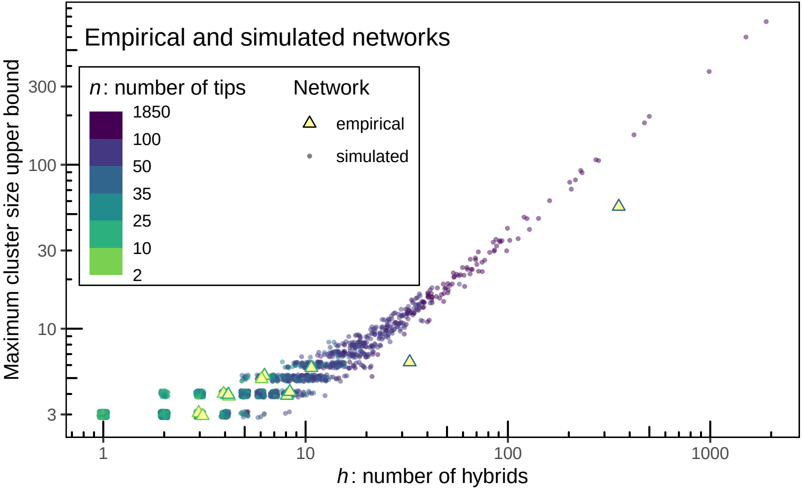
Linear Gaussian

- Only depends on $\mathbb{E}[\mathbf{X}_v \mid \mathbf{Y}]$ and $\mathbb{V}\mathrm{ar}[\mathbf{X}_v \mid \mathbf{Y}]$
- sub-product of BP
- links with autodiff ?

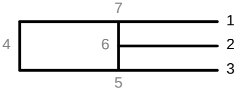
Simple Brownian Motion: analytical estimator formulas

- for the multivariate BM and phylogenetic regression
- network equivalent to Ho and Ané (2014)

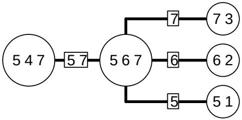
When is BP efficient ?



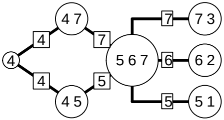
Cluster Graph



Phylogenetic Network



Clique Tree



Cluster Graph

Cluster graph:

- Nodes: clusters C_i (family preserving)
- Edges: Sepsets $S_{ij} \subseteq C_i \cap C_j$
- Running intersection: each variable defines a tree

Construction:

- Bethe cluster graph, join graph, ...
- Trade-off complexity / accuracy

Loopy BP

Conclusion

General framework for trait evolution on networks

- Ecology: continuous traits
- Virology: phylogeography
- Admixture graph inference

Implementation

- julia package
- linear time estimators for the BM
- work on API and integration within `julia`

Perspectives

- loopy BP vs BP: network structures ?
- Deal with degeneracy: traits at tips
- Beyond Gaussian: discrete traits

B. Teo, P. Bastide, C. Ané (2024+), Leveraging graphical model techniques to study evolution on phylogenetic networks. *in prep*.

Bibliography I

Cappé, Moulines, Rydén. 2005. Springer Series in Statistics. New York, NY: Springer New York.

Felsenstein. 1985. *The American Naturalist*. 125:1–15.

Hansen. 1997. *Evolution*. 51:1341.

Harmon, Losos, Davies, et al. 2010. *Evolution*. 64:2385–2396.

Ho, Ané. 2014. *Methods in Ecology and Evolution*. 5:1133–1146.

Koller, Friedman. 2009. The MIT Press.

Müller, Kistler, Bedford. 2022. *Nature Communications*. 13:4186.

Nielsen, Vaughn, Leppälä, et al. 2023. *PLOS Genetics*. 19:e1010410.

Pybus, Suchard, Lemey, et al. 2012. *Proceedings of the National Academy of Sciences*. 109:15066–15071.

Teo, Bastide, Ané. 2024+. *in prep.* .

Teo, Rose, Bastide, et al. 2023. *Bulletin of the Society of Systematic Biologists*. 3:1–29.

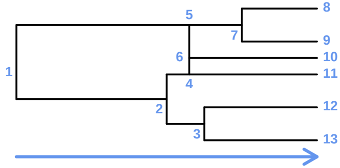
Photo Credits
- Polemonium eddyense: Gail (iNaturalist user ribes2018), CC BY 4.0 https://creativecommons.org/licenses/by/4.0, via Wikimedia Commons
- Polemonium carneum: Adamschneider, CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0, via Wikimedia Commons
- Polemonium reptens: Photo by and (c)2007 Derek Ramsey (Ram-Man), GFDL 1.2 http://www.gnu.org/licenses/old-licenses/fdl-1.2.html, via Wikimedia Commons
- Patrick Giraud (edited to fix white balance), CC BY-SA 3.0 http://creativecommons.org/licenses/by-sa/3.0/, via Wikimedia Commons
- Clément Bardot, CC BY-SA 4.0 https://creativecommons.org/licenses/by-sa/4.0, via Wikimedia Commons

Thank you for listening

Institut Montpelliérain Alexander Grothendieck

Appendices

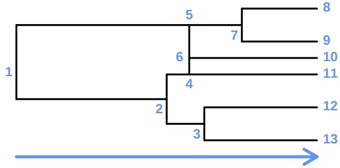
Pre-Order Computation



Pre-order

Root: $C_{11} = 0$

Pre-Order Computation



Pre-order

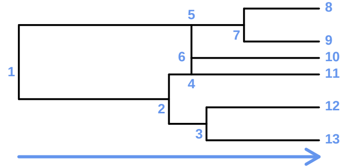
Root: $C_{11} = 0$

Tree node i with parent a

$$\begin{cases} C_{ij} = C_{aj} & j < i \\ C_{ii} = C_{aa} + \ell_a \end{cases}$$

$$X_i = X_a + \epsilon_a \qquad \epsilon_a \sim \mathcal{N}(0, \ell_a)$$

Pre-Order Computation

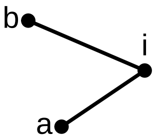


Pre-order

Root: $C_{11} = 0$

Tree node i with parent a

$$\begin{cases} C_{ij} = C_{aj} & j < i \\ C_{ii} = C_{aa} + \ell_a \end{cases}$$

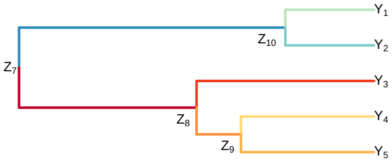


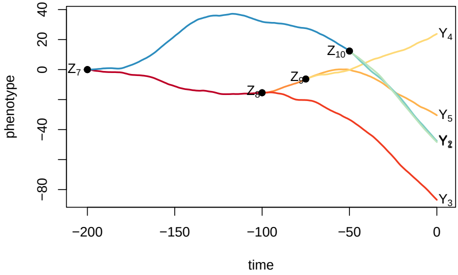
$$X_i = \gamma_a(X_a + \epsilon_a) \quad \epsilon_a \sim \mathcal{N}(0, \ell_a)$$
$$+ \gamma_b(X_b + \epsilon_b) \quad \epsilon_b \sim \mathcal{N}(0, \ell_b)$$

Hybrid node i with parents a and b

$$\begin{cases} C_{ij} = \gamma_a C_{aj} + \gamma_b C_{bj} & j < i \\ C_{ii} = \gamma_a^2(C_{aa} + \ell_a) \\ \qquad + \gamma_b^2(C_{bb} + \ell_b) \\ \qquad + 2\gamma_a\gamma_b C_{ab} \end{cases}$$

Integrated Brownian Motion

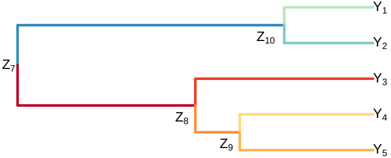


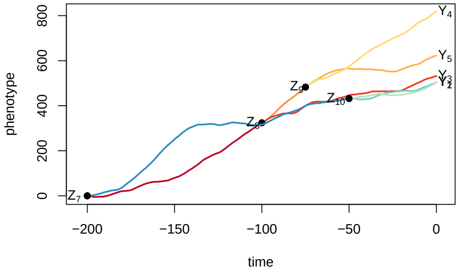


IBM heredity

$$\begin{pmatrix} V_i \\ X_i \end{pmatrix} \Bigg| \begin{pmatrix} V_{\mathrm{pa}(i)} \\ X_{\mathrm{pa}(i)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 & 0 \\ \ell_i & 1 \end{pmatrix} \begin{pmatrix} V_{\mathrm{pa}(i)} \\ X_{\mathrm{pa}(i)} \end{pmatrix} ; \sigma^2 \begin{pmatrix} \ell_i & \ell_i^2/2 \\ \ell_i^2/2 & \ell_i^3/3 \end{pmatrix} \right)$$

Integrated OU





IOU heredity

back

$$\begin{pmatrix} V_i \\ X_i \end{pmatrix} \Big| \begin{pmatrix} V_{\mathrm{pa}(i)} \\ X_{\mathrm{pa}(i)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} e^{-\alpha\ell_i} & 0 \\ (1-e^{-\alpha\ell_i})/\alpha & 1 \end{pmatrix} \begin{pmatrix} V_{\mathrm{pa}(i)} \\ X_{\mathrm{pa}(i)} \end{pmatrix} + \begin{pmatrix} (1-e^{-\alpha\ell_i})\beta \\ (\ell_i - (1-e^{-\alpha\ell_i})/\alpha)\beta \end{pmatrix}; \right.$$

$$\left. \frac{\sigma^2}{2\alpha} \begin{pmatrix} (1-e^{-2\alpha\ell_i}) & \frac{(1-e^{-\alpha\ell_i})^2}{\alpha} \\ \frac{(1-e^{-\alpha\ell_i})^2}{\alpha} & \frac{2\ell_i}{\alpha} - 4\frac{(1-e^{-\alpha\ell_i})}{\alpha^2} + \frac{(1-e^{-2\alpha\ell_i})}{\alpha^2} \end{pmatrix} \right)$$

Loopy BP

Loopy BP

- Choose a *schedule*
- Pass messages until convergence (not guaranteed)
- $q = \frac{\prod_{C_i \in \mathcal{V}^*} \beta_i}{\prod_{\{C_i, C_j\} \in \mathcal{E}^*} \mu_{i,j}} \approx p(X_v, v \notin V_{obs} | X_v, v \in V_{obs})$

First approximation: ELBO

$$\log p(X_v, v \in V_{obs}) \geq \mathbb{E}_q[\log p_\theta(X_v, v \in V)] - \mathbb{E}_q[\log q(X_v, v \notin V_{obs})]$$

Energy functional

$$F(p_\theta, q) = \sum_{C_i \in \mathcal{V}^*} \mathbb{E}_q(\log \psi_i) + \sum_{C_i \in \mathcal{V}^*} \mathbb{E}_q[-\log \beta_i] - \sum_{\{C_i, C_j\} \in \mathcal{E}^*} \mathbb{E}_q[-\log \mu_{i,j}]$$

Loopy BP

Second approximation: localization

$$\mathbb{E}_q(\log \psi_i) \approx \mathbb{E}_{\beta_i}(\log \psi_i)$$

Factored Energy functional

$$\tilde{F}(p_\theta, q) = \sum_{C_i \in \mathcal{V}^*} \mathbb{E}_{\beta_i}(\log \psi_i) + \sum_{C_i \in \mathcal{V}^*} \mathbb{E}_{\beta_i}[-\log \beta_i] - \sum_{\{C_i, C_j\} \in \mathcal{E}^*} \mathbb{E}_{\mu_{i,j}}[-\log \mu_{i,j}] \,.$$

Loopy BP at calibration

$$\log p(X_v, v \in V_{obs}) \approx \tilde{F}(p_\theta, q)$$

back