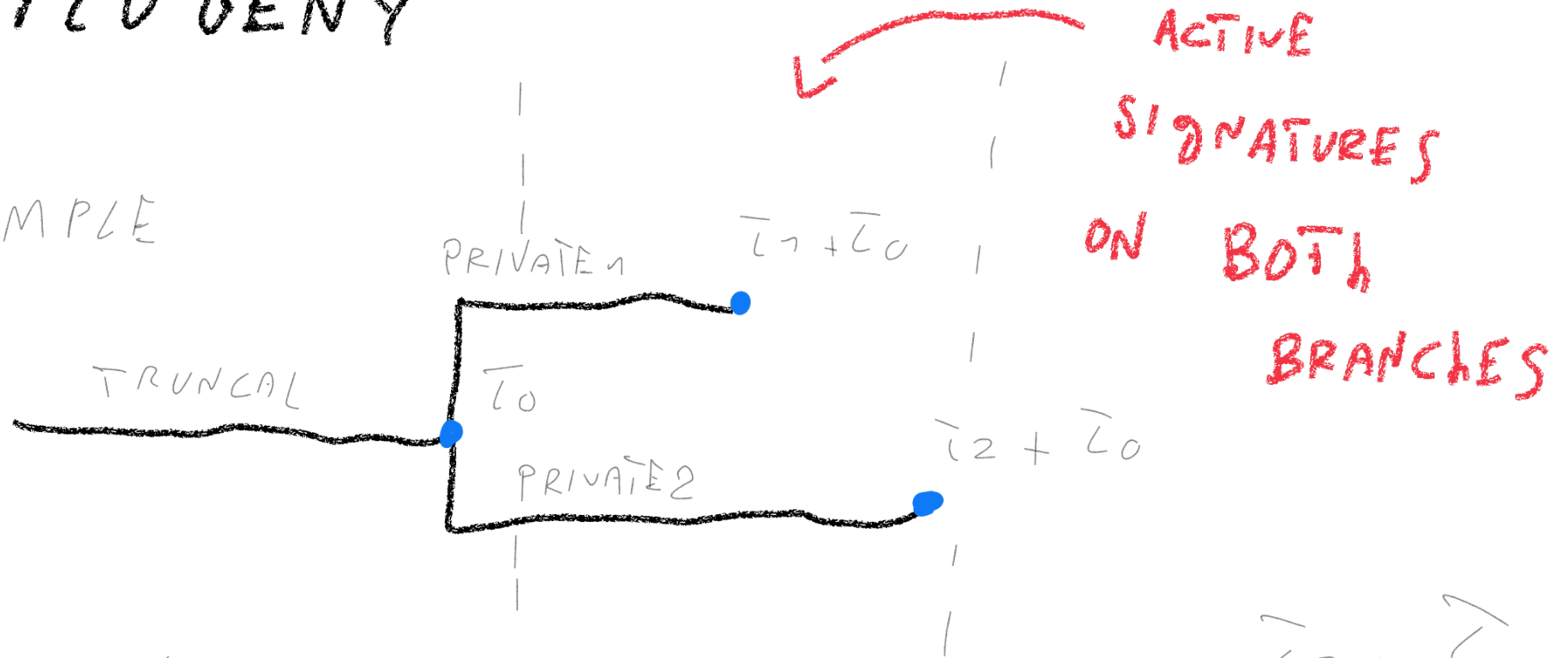


SIGNATURE DE CONVOLUTION BASED ON PHYLOGENY

EXAMPLE



$T_0 = \#$ TRUNCAL MUTATIONS

$T_1 = \#$ PRIVATE MUTATIONS BRANCH 1

$T_2 = \#$ PRIVATE MUTATIONS BRANCH 2

$$T_2 > T_1$$

INPUT MATRIX

$$M = \begin{pmatrix} M_{\text{TRUNCAL}} \\ M_{\text{PRIVATE 1}} \\ M_{\text{PRIVATE 2}} \end{pmatrix}$$

3 X 96 MATRIX
↓
NUM. SAMPLES NUM. CONTEXTS

$$M \sim \text{POISSON}(\alpha, \beta)$$

↓
ACTIVITIES

→ SIGNATURE
PROFILES

$$\beta = \begin{pmatrix} \beta_{SBS9} \\ \beta_{SBS5} \\ \beta_{chemo} \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \begin{matrix} \left. \vphantom{\begin{pmatrix} \beta_{SBS9} \\ \beta_{SBS5} \end{pmatrix}} \right\} 2 \\ \left. \vphantom{\begin{pmatrix} \beta_{chemo} \\ \vdots \end{pmatrix}} \right\} K \end{matrix}$$

$K+2 \times 96$
MATRIX

SBS9, SBS5
/

AGING

SIGNATURES:
FIXED

PRIORS

SIGNATURE PROFILES ARE PROBABILITY DISTRIBUTIONS
OVER 96 CONTEXTS

WE USE DIRICHLET PRIORS:

$$\beta_{\text{chemo}} \sim \text{DIRICHLET}(\rho_{\text{chemo}})$$

$$\text{we choose } \rho_{\text{chemo}} = \left(\frac{1}{96}, \dots, \frac{1}{96} \right)$$

$$Z = \begin{pmatrix} Z_{\text{TRUNCAL}} \\ Z_{P21} \\ Z_{P22} \end{pmatrix}$$

$3 \times K+2$ MATRIX

RECAST AS $3K+6$ VECTOR

DEFINE $X = \log(Z)$

$X \sim \text{MULTIVARIATE NORMAL} (0, \Sigma)$

\sum PROVIDE THE CORRELATION STRUCTURE

POSSIBLE DESIGN

$$\sum_{i,i} = 1$$

WE CORRELATE THE ACTIVITIES OF DIFFERENT SAMPLE
FOR THE SAME SIGNATURE

$$\text{GIVEN } (J-i) \% K+2 = 0, \quad i \neq J$$

\downarrow
MODULO

WE WRITE

$$m_2 = \left[\frac{i}{k+2} \right]$$

$$m_2 = \left[\frac{j}{k+2} \right]$$

$$[] = \text{integer part}$$

DEFINE:

OVERLAP MATRIX

ENCODING THE LENGTH

OF THE BRANCH OVERLAPS

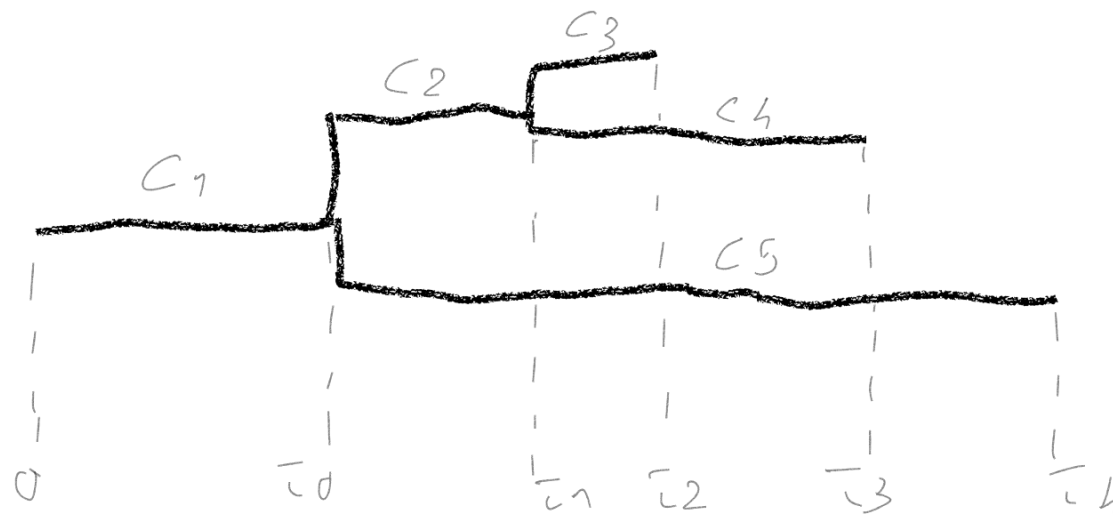
IN TERMS OF NUMBER

OF MUTATIONS

$$\text{OVERLAP} = \begin{pmatrix} \tau_0 & 0 & 0 \\ 0 & \tau_1 & \tau_1 \\ 0 & \tau_1 & \tau_2 \end{pmatrix}$$

$$\Sigma_{i,j} = \exp \left(- \frac{1}{\text{OVERLAP}(m(i), m(j))} \right)$$

GENERALIZE TO ARBITRARY PHYLOGENY



DIPLOID

MUTATIONS

TIMES =

MUTATIONAL
DISTANCES

INPUT
DATA

$$M = \begin{pmatrix} \vdots \\ c_i \\ \vdots \end{pmatrix}$$

N X 96 MATRIX

N = # branches

OVERLAP
MATRIX

$$O_{ij} = \min(\bar{t}_{end_i}, \bar{t}_{end_j}) - \max(\bar{t}_{start_i}, \bar{t}_{start_j})$$

\bar{t}_{start_1}

\bar{t}_{end_1}



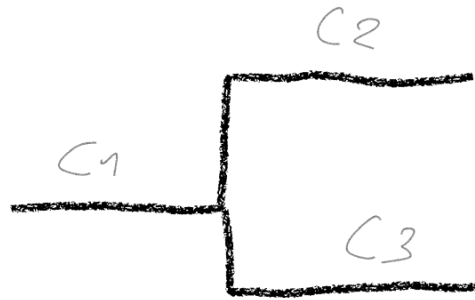
\bar{t}_{end_2}



\bar{t}_{start_2}

$$O = \bar{t}_{end_1} - \bar{t}_{start_2}$$

ADJACENCY MATRIX



$$A_{ij} = \begin{cases} 1 & \text{IF TWO BRANCHES ARE CONTIGUOUS} \\ 0 & \text{OTHERWISE} \end{cases}$$

$$A_{12} = A_{21} = 1$$

$$A_{23} = 0$$

$$A_{13} = A_{31} = 1$$

$$M = \alpha \beta$$

$$\beta = \begin{pmatrix} \beta_{k1} \\ \beta_{k2} \end{pmatrix} \quad \begin{matrix} K \times 26 \\ \text{MATRIX} \end{matrix}$$

$$\alpha \quad N \times K \quad \text{MATRIX}$$

ACTIVITIES

$$K = K_1 + K_2$$

\downarrow \downarrow \searrow
 TOT KNOWN UNKNOWN
 SIGNATURES

$\beta \sim \text{Dirichlet} \left(\frac{1}{96}, \dots, \frac{1}{96} \right)$

$$\lambda = \Theta \cdot P$$

$$\Theta = \begin{pmatrix} \Theta_1 & & & \\ & \ddots & & \\ & & \bigcirc & \\ & \bigcirc & & \ddots \\ & & & & \Theta_N \end{pmatrix}$$

Θ_i : MUTATIONAL
DISTANCE
BRANCH C_i

P = PROBABILITY MATRIX

$$P = \begin{pmatrix} \overrightarrow{P_1} \\ \vdots \\ \overrightarrow{P_N} \end{pmatrix}$$

$\overrightarrow{P_i}$
↓

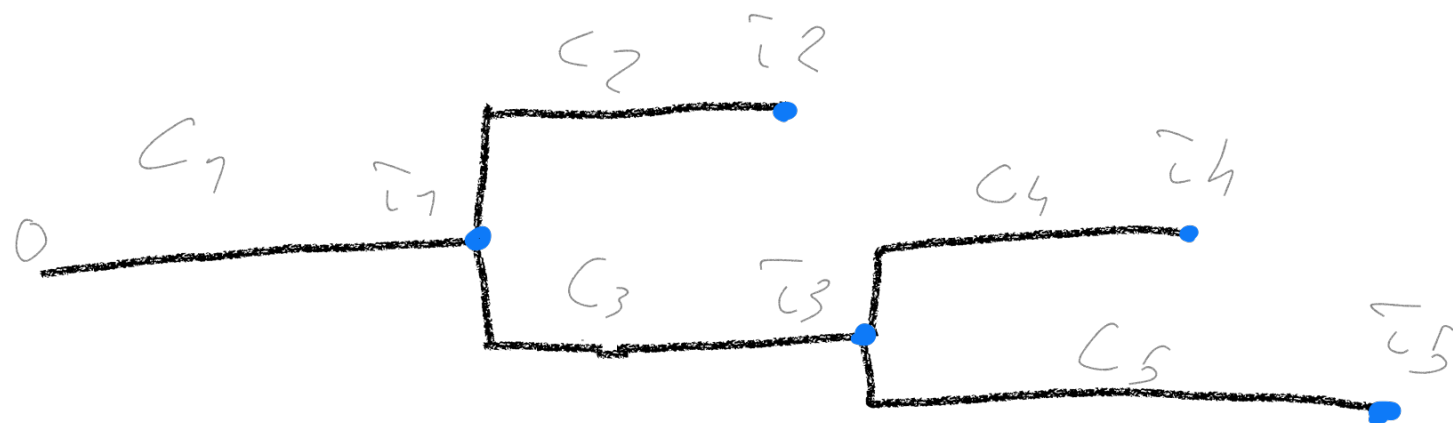
K VECTOR

$$P_{ij} > 0$$

$$\sum_j P_{ij} = 1$$

C_i BRANCH

TRANSFER LEARNING USING THE PHYLOGENY



$C_1 \rightarrow C_2$

$C_3 \rightarrow C_4$

$C_1 \rightarrow C_3$

$C_3 \rightarrow C_5$

$C_2 \leftarrow C_3$

$C_4 \leftarrow C_5$

DEFINE A GRAPH:

C_i = VERTICES

TWO ADJACENCY MATRICES

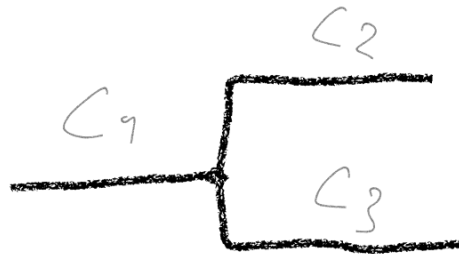
A_T A_S
↓ ↓
TEMPORAL SPATIAL

A_T = BINARY

SPECIFY TEMPORAL RELATIONS

$A_{ii} = 1$

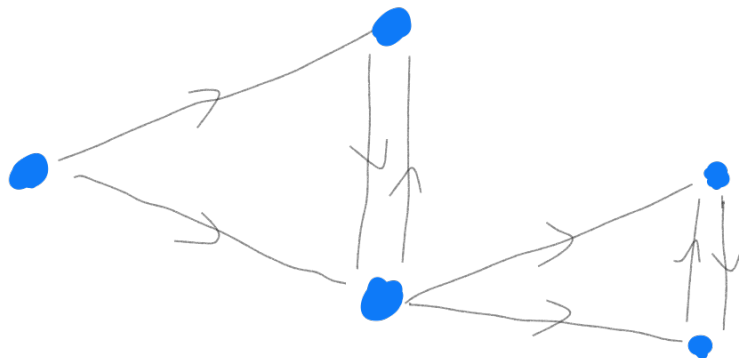
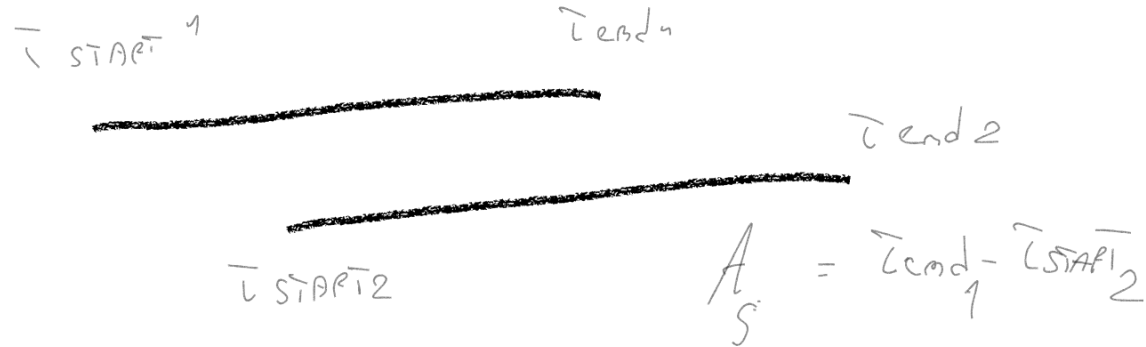
$A_{ij} = 0, 1$



$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

OVERLAP MATRIX

$$(A_S)_{ij} = \min(\bar{t}_{end_i}, \bar{t}_{end_j}) - \max(\bar{t}_{start_i}, \bar{t}_{start_j})$$

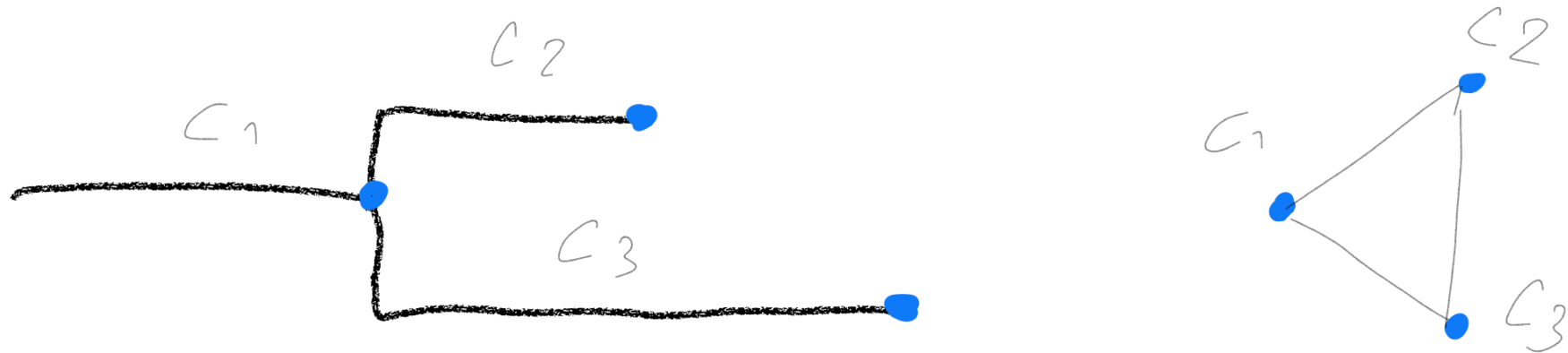


TEMPORAL
edge



OVERLAP
edge

DEFINE A MAP GRAPH \Leftrightarrow PHYLOGENY



ASSOCIATE BINARY ADJACENCY MATRIX A

where $A_{ij} = \begin{cases} 1 & \text{IF TWO NODES ARE} \\ & \text{PHYLOGENETIC RELATED OR} \\ & \text{OVERLAPS} \\ 0 & \text{OTHERWISE} \end{cases}$

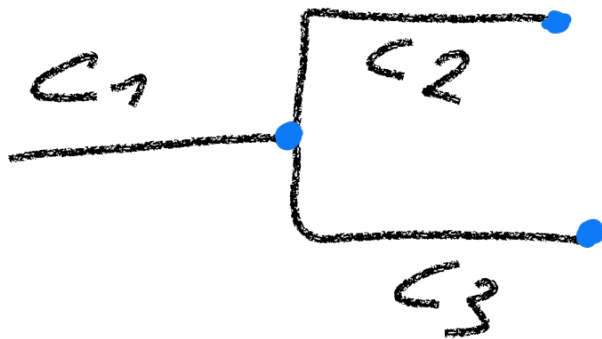
FIRST RUN : INDEPENDENT INFERENCE

TRANSFER 2 USING POSITIVE TRANSFER

SCORES (COSINE SIMILARITIES)

$$\beta = \gamma_{\text{LOCAL}}$$

ex



$$\alpha_3' = s_1 \alpha_1 + s_2 \alpha_2 + s_3 \alpha_3$$

↓
NEW PRIOR OR
INITIALIZING POINT

$$\lambda = \begin{pmatrix} \Theta_1 & & 0 \\ & \ddots & \\ 0 & & \Theta_N \end{pmatrix} \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_{\text{known}} \\ \beta_{\text{unknown}} \end{pmatrix} = \eta_{\text{local}}$$

$Q_i = \text{TOT MUTATIONS SAMPLE } i$
estimates FROM THE DATA

STEP 1 : INFERENCE WITH FLAT PRIORS

$$\beta_0 \sim \text{DIRICHLET}(d_1, \dots, d_K)$$

$$d_j = \frac{1}{96}$$

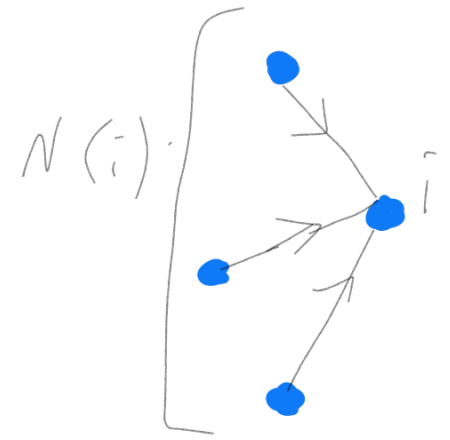
$$z_j = \frac{1}{K}$$

$$p_i \sim \text{DIRICHLET}(z_1, \dots, z_K)$$

DO MAP INFERENCE

PRIOR UPDATES :

$$\vec{d}_{step_{i+1}} = \vec{\beta}_{step_i}$$



$$\vec{z}_{i,j+1} = \sum_{l \in N(i)} c_l \vec{p}_{l,j} + c_i \vec{p}_{i,j}$$

\downarrow node \downarrow step \downarrow neighbours

c_l = TRANSFER COEFF

TRANSFER coeff and CONV. COND

$$C_2 = \lambda S_c(M_i, \Theta_i, P_2, \beta)$$

OPTIMIZATION
COEFF

↓
CROSS SAMPLE COS. SIMILARITY

$$C_i = (1 - \lambda) S_c(M_i, \Theta_i, P_i, \beta)$$

↓
SELF RECONSTRUCTION COS. SIMILARITY

UPDATES UNTIL $\|P_{i,j+1} - P_{i,j}\|^2 < \epsilon \quad \forall i$
given ϵ

FIT For a range of λ VALUES

$\lambda = 0, 0.1, \dots$ and choose the best

RECONSTRUCTION

Choose $\lambda \in [0, 1]$

WE CAN NORMALIZE THE T_2 COEFF TO 1

$$w_{2,i} = \frac{S_c(m_i, \theta_i, \rho, \beta)}{\sum_j S_c(m_i, \theta_i, \rho_j, \beta)}$$

$$C^i_l = \lambda W^i_l \quad i \neq l$$

$$C^i_i = (1 - \lambda) W^i_i \quad i = l$$

λ = hyperparameter

$$\sum_l W^i_l = 1 \quad \forall i$$