Publication venue: **Data Mining and Knowledge Discovery**

Research Article
30 October 2015

Supplementary Material: ClusPath: A Temporal-driven Clustering Solution to Inferring Typical Evolution Paths

Marian-Andrei Rizoiu [1,*], Julien Velcin [2], Stéphane Bonnevay[3] and Stéphane Lallich [2]

[1] *NICTA Research Lab & Australian National University, Canberra, Australia*
[2] *Laboratory ERIC, University Lumière Lyon 2, Lyon, France*
[3] *Laboratory ERIC, University Claude Bernard Lyon 1, Lyon, France*

Correspondence*:
Marian-Andrei Rizoiu
NICTA Research Lab, 7 London Circuit, Canberra, Australia,
Marian-Andrei.Rizoiu@nicta.com.au

The purpose of this document is to detail information, which would be i) cumbersome for the main article and ii) of limited interest for the main scientific message presented in the main article. This document does not introduce new notions and the information presented here-after is not necessary for the comprehension of the main article. We present it for completeness and reproducibility reasons.

We produce hereafter, the complete calculations of the prototypes update formulas, as well as the formulas for updating the adjacency matrix and runtimes for the algorithms.

1 INFERRING THE PROTOTYPES UPDATE FORMULAS

The objective function $\mathcal{J}$ that needs to be minimized using a gradient descend K-Means-like framework is (defined in Equation 8 in the Main Text (MT)):

$$\begin{aligned}
\mathcal{J} &= \lambda_1 T_1 + \lambda_2 T_2 + \lambda_3 T_3 = \\
&= \lambda_1 \sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \left(||x_i - \mu_p||_{TA} + \sum_{\substack{x_k \in \mathcal{C}_q \\ q \neq p \\ x_i^\phi = x_k^\phi}}^{x_i^t < x_k^t} \beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} \left(1 - a_{p,q}^2\right) \right) + \\
&+ \lambda_2 \sum_{\mu_p \in \mathcal{M}} \sum_{\substack{\mu_q \in \mathcal{M} \\ p \neq q}} a_{p,q}^2 ||\mu_p - \mu_q||_{TA} + \\
&+ \lambda_3 \sum_{\mu_p \in \mathcal{M}} \sum_{\substack{\mu_q \in \mathcal{M} \\ p \neq q}} a_{p,q}^2 inter_\phi^2(\mathcal{C}_p, \mathcal{C}_q) \quad ,
\end{aligned} \tag{1}$$

and the temporal-aware dissimilarity measure is defined as (in Equation 1 in MT):

$$||x_i - x_j||_{TA} = 1 - \left(1 - \gamma_d \frac{||x_i^d - x_j^d||^2}{\Delta d_{max}^2}\right)\left(1 - \gamma_t \frac{||x_i^t - x_j^t||^2}{\Delta t_{max}^2}\right) \quad , \tag{2}$$

where the weight of the descriptive component and, respectively, the temporal component, are controlled by the parameter $\alpha \in [-1, 1]$:

$$\gamma_d = \begin{cases} 1+\alpha, & \text{if } \alpha \in [-1,0] \\ 1, & \text{if } \alpha \in (0,1] \end{cases} \; ; \; \gamma_t = \begin{cases} 1, & \text{if } \alpha \in [-1,0] \\ 1-\alpha, & \text{if } \alpha \in (0,1] \end{cases} \quad .$$

As stated in the main article, calculating the update formulas for the prototypes boils down to recomputing each of the descriptive components (μ_j^d) and temporal component (μ_j^t) of the prototypes. More precisely, we are searching for the fixed point, by calculating the derivative of the function $\mathcal{J}$ of the variables μ_j^d and μ_j^t. Therefore, the following system of equations needs to be solved:

$$\frac{\partial \mathcal{J}}{\partial \mu_j^d} = 0 \; ; \; \frac{\partial \mathcal{J}}{\partial \mu_j^t} = 0$$

We exemplify the calculation of the derivative of $\mathcal{J}$ of the variable μ_j^d. From Equation 1 we obtain:

$$\frac{\partial \mathcal{J}}{\partial \mu_j^d} = \lambda_1 \frac{\partial T_1}{\partial \mu_j^d} + \lambda_2 \frac{\partial T_2}{\partial \mu_j^d} + \lambda_3 \frac{\partial T_3}{\partial \mu_j^d} \tag{3}$$

Knowing that the $inter_\phi$ function from the term T_3 is defined as:

$$inter_\phi(\mathcal{C}_p, \mathcal{C}_q) = 1 - \frac{|\{\phi_l \in \Phi | \mathcal{C}_p \xrightarrow{\phi_l} \mathcal{C}_q\}|}{|\Phi|} \quad ,$$

and, therefore, it is independent of μ_j^d, the following holds:

$$\frac{\partial T_3}{\partial \mu_j^d} = 0, \forall \mu_j^d \tag{4}$$

We augment the definition of the first two terms, by introducing the temporal-aware dissimilarity measure (in Equation 2) into the formula of the objective function (Equation 1):

$$T_1 = \sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \left(1 - \left(1 - \gamma_d \frac{||x_i^d - \mu_p^d||^2}{\Delta d_{max}^2}\right)\left(1 - \gamma_t \frac{||x_i^t - \mu_p^t||^2}{\Delta t_{max}^2}\right) + \sum_{\substack{x_i^t < x_k^t \\ x_k \in \mathcal{C}_q \\ q \neq p \\ x_i^\phi = x_k^\phi}} \beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} (1 - a_{p,q}^2) \right)$$

$$T_2 = \sum_{\mu_p \in \mathcal{M}} \sum_{\substack{\mu_q \in \mathcal{M} \\ p \neq q}} a_{p,q}^2 \left(1 - \left(1 - \gamma_d \frac{||\mu_p^d - \mu_q^d||^2}{\Delta d_{max}^2}\right)\left(1 - \gamma_t \frac{||\mu_p^t - \mu_q^t||^2}{\Delta t_{max}^2}\right)\right) \quad .$$

We calculate the derivative of T_1. We observe that the penalty term $\sum_{\substack{x_i^t < x_k^t \\ x_k \in \mathcal{C}_q \\ q \neq p \\ x_i^\phi = x_k^\phi}} \beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} \left(1 - a_{p,q}^2\right)$ is independent of μ_j^d and, therefore, it is nullified when calculating the derivative:

$$\frac{\partial T_1}{\partial \mu_j^d} = \frac{\partial}{\partial \mu_j^d} \left(\sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \left(1 - \left(1 - \gamma_d \frac{||x_i^d - \mu_p^d||^2}{\Delta d_{max}^2} \right) \left(1 - \gamma_t \frac{||x_i^t - \mu_p^t||^2}{\Delta t_{max}^2} \right) \right) \right) \tag{5}$$

Considering that the double sum iterates over all the observations:

$$\sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} 1 = |\mathcal{X}|$$

Equation 5 becomes:

$$\begin{aligned}
\frac{\partial T_1}{\partial \mu_j^d} &= \frac{\partial |\mathcal{X}|}{\partial \mu_j^d} - \frac{\partial}{\partial \mu_j^d} \left(\sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \left(1 - \gamma_d \frac{||x_i^d - \mu_p^d||^2}{\Delta d_{max}^2} \right) \left(1 - \gamma_t \frac{||x_i^t - \mu_p^t||^2}{\Delta t_{max}^2} \right) \right) = \\
&= - \frac{\partial}{\partial \mu_j^d} \left(\sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \left(1 - \gamma_d \frac{||x_i^d - \mu_p^d||^2}{\Delta d_{max}^2} \right) \left(1 - \gamma_t \frac{||x_i^t - \mu_p^t||^2}{\Delta t_{max}^2} \right) \right) = \\
&= - \sum_{x_i \in \mathcal{C}_j} \frac{\partial}{\partial \mu_j^d} \left(1 - \gamma_d \frac{||x_i^d - \mu_j^d||^2}{\Delta d_{max}^2} \right) \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) = \\
&= - \frac{2\gamma_d}{\Delta d_{max}^2} \sum_{x_i \in \mathcal{C}_j} \left(x_i^d - \mu_j^d \right) \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) \quad .
\end{aligned} \tag{6}$$

To ease the calculation of the derivative of T_2, we make the following notation:

$$s_{pq} = \left(1 - \gamma_t \frac{||\mu_p^t - \mu_q^t||^2}{\Delta t_{max}^2} \right) \quad . \tag{7}$$

Therefore:

$$\frac{\partial T_2}{\partial \mu_j^d} = \frac{\partial}{\partial \mu_j^d} \sum_{\mu_p \in \mathcal{M}} \sum_{\substack{\mu_q \in \mathcal{M} \\ p \neq q}} a_{p,q}^2 \left(1 - \left(1 - \gamma_d \frac{||\mu_p^d - \mu_q^d||^2}{\Delta d_{max}^2} \right) s_{pq} \right) \quad . \tag{8}$$

The term μ_j^d can appear twice: i) once in the first sum $\sum_{\mu_p \in \mathcal{M}}$, for $p = j$ and ii) once in the second sum $\sum_{\substack{\mu_q \in \mathcal{M} \\ p \neq q}}$, for $q = j$. Equation 8 can be rewritten as:

$$\frac{\partial T_2}{\partial \mu_j^d} = \sum_{\substack{\mu_q \in \mathcal{M} \\ q \neq j}} \frac{\partial}{\partial \mu_j^d} \left(a_{j,q}^2 \left(1 - \left(1 - \gamma_d \frac{||\mu_j^d - \mu_q^d||^2}{\Delta d_{max}^2} \right) s_{jq} \right) \right)$$

$$+ \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \frac{\partial}{\partial \mu_j^d} \left(a_{p,j}^2 \left(1 - \left(1 - \gamma_d \frac{||\mu_p^d - \mu_j^d||^2}{\Delta d_{max}^2} \right) s_{pj} \right) \right) =$$

$$= \sum_{\substack{\mu_q \in \mathcal{M} \\ q \neq j}} \frac{\partial}{\partial \mu_j^d} \left(\frac{a_{j,q}^2 \gamma_d}{\Delta d_{max}^2} ||\mu_j^d - \mu_q^d||^2 s_{jq} \right) + \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \frac{\partial}{\partial \mu_j^d} \left(\frac{a_{p,j}^2 \gamma_d}{\Delta d_{max}^2} ||\mu_p^d - \mu_j^d||^2 s_{pj} \right) =$$

$$= \frac{2\gamma_d}{\Delta d_{max}^2} \left[\sum_{\substack{\mu_q \in \mathcal{M} \\ q \neq j}} a_{j,q}^2 (\mu_j - \mu_q) s_{jq} - \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} a_{p,j}^2 (\mu_p - \mu_j) s_{pj} \right] . \tag{9}$$

We re-note the sum indexes so that $p = q$ and we observe that s_{pq} defined in Equation 7 is symmetric: $s_{pq} = s_{qp}$. Consequently, we re-write Equation 9 as:

$$\frac{\partial T_2}{\partial \mu_j^d} = \frac{2\gamma_d}{\Delta d_{max}^2} \left[\mu_j^d \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} a_{j,p}^2 s_{jp} - \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d a_{j,p}^2 s_{jp} + \mu_j^d \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} a_{p,j}^2 s_{pj} - \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d a_{p,j}^2 s_{pj} \right] =$$

$$= \frac{2\gamma_d}{\Delta d_{max}^2} \left[\mu_j^d \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} s_{pj} (a_{j,p}^2 + a_{p,j}^2) - \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d s_{pj} (a_{j,p}^2 + a_{p,j}^2) \right] . \tag{10}$$

By introducing Equations 4, 6 and 10 into Equation 3, we can calculate the formula of the derivative of function $\mathcal{J}$ of variable μ_j^d and we calculate the fixed point:

$$\frac{\partial \mathcal{J}}{\partial \mu_j^d} = 0 \Leftrightarrow \frac{2\gamma_d}{\Delta d_{max}^2} \left[\lambda_1 \mu_j^d \sum_{x_i \in \mathcal{C}_j} \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) - \lambda_1 \sum_{x_i \in \mathcal{C}_j} x_i^d \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) + \right.$$

$$\left. + \lambda_2 \mu_j^d \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} s_{pj} (a_{j,p}^2 + a_{p,j}^2) - \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d s_{pj} (a_{j,p}^2 + a_{p,j}^2) \right] = 0$$

$$\Leftrightarrow \mu_j^d \left[\lambda_1 \sum_{x_i \in \mathcal{C}_j} \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} s_{pj} (a_{j,p}^2 + a_{p,j}^2) \right] =$$

$$= \lambda_1 \sum_{x_i \in \mathcal{C}_j} x_i^d \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2} \right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d s_{pj} (a_{j,p}^2 + a_{p,j}^2)$$

$$\Leftrightarrow \mu_j^d = \frac{\lambda_1 \sum_{x_i \in \mathcal{C}_j} x_i^d \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d s_{pj} (a_{j,p}^2 + a_{p,j}^2)}{\lambda_1 \sum_{x_i \in \mathcal{C}_j} \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} s_{pj} (a_{j,p}^2 + a_{p,j}^2)} \quad . \tag{11}$$

By introducing the notation for s_{pq} (defined in Equation 7) back into Equation 11, we obtain the final centroid update formula for the descriptive component:

$$\mu_j^d = \frac{\lambda_1 \sum_{x_i \in \mathcal{C}_j} x_i^d \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^d \left(1 - \gamma_t \frac{||\mu_p^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) (a_{j,p}^2 + a_{p,j}^2)}{\lambda_1 \sum_{x_i \in \mathcal{C}_j} \left(1 - \gamma_t \frac{||x_i^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \left(1 - \gamma_t \frac{||\mu_p^t - \mu_j^t||^2}{\Delta t_{max}^2}\right) (a_{j,p}^2 + a_{p,j}^2)} \quad .$$

Similarly, we can deduce the update formula for the temporal component of prototypes:

$$\mu_j^t = \frac{\lambda_1 \sum_{x_i \in \mathcal{C}_j} x_i^t \left(1 - \gamma_d \frac{||x_i^d - \mu_j^d||^2}{\Delta d_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \mu_p^t \left(1 - \gamma_d \frac{||\mu_p^d - \mu_j^d||^2}{\Delta d_{max}^2}\right) (a_{j,p}^2 + a_{p,j}^2)}{\lambda_1 \sum_{x_i \in \mathcal{C}_j} \left(1 - \gamma_d \frac{||x_i^d - \mu_j^d||^2}{\Delta d_{max}^2}\right) + \lambda_2 \sum_{\substack{\mu_p \in \mathcal{M} \\ p \neq j}} \left(1 - \gamma_d \frac{||\mu_p^d - \mu_j^d||^2}{\Delta d_{max}^2}\right) (a_{j,p}^2 + a_{p,j}^2)} \quad .$$

2 UPDATING THE ADJACENCY MATRIX

The objective function $\mathcal{J}$, in Equation 1, can be trivially minimized by setting $a_{p,q} = 0, \forall p, q \in [1, k]$. To avoid this situation, an additional constraint is imposed on the 1-norm of the adjacency matrix (in Equation 9 in MT):

$$||A||_1 = 1 \Leftrightarrow \sum_{p=1}^{k} \sum_{q=1}^{k} a_{p,q} = 1 \quad . \tag{12}$$

The purpose of updating the adjacency matrix is to find the best adjacency matrix A^* which minimizes the objective function $\mathcal{J}$:

$$A^* = \arg\min_A \mathcal{J} \quad .$$

A classical strategy for finding the local maxima and minima of a function subject to equality constraints is method of Lagrange multipliers. This constructs a new problem:

$$A^* = \arg\min_A \mathcal{J}^*, \text{ where } \mathcal{J}^* = \mathcal{J} - \lambda \left(\sum_{p=1}^{k} \sum_{q=1}^{k} a_{p,q} - 1 \right) \quad .$$

By calculating the derivative of $\mathcal{J}^*$ of each $a_{r,s}$, we obtain the formula of the local optimum:

$$\frac{\partial \mathcal{J}^*}{\partial a_{r,s}} = \frac{\partial}{\partial a_{r,s}} \left(\mathcal{J} - \lambda \left(\sum_{p=1}^{k} \sum_{q=1}^{k} a_{p,q} - 1 \right) \right) = \frac{\partial \mathcal{J}}{\partial a_{r,s}} - \lambda \frac{\partial}{\partial a_{r,s}} \left(\sum_{p=1}^{k} \sum_{q=1}^{k} a_{p,q} - 1 \right) = \frac{\partial \mathcal{J}}{\partial a_{r,s}} - \lambda \quad . \tag{13}$$

All three terms of the objective function are dependent on $a_{r,s}$, therefore:

$$\frac{\partial \mathcal{J}}{\partial a_{r,s}} = \lambda_1 \frac{\partial T_1}{\partial a_{r,s}} + \lambda_2 \frac{\partial T_2}{\partial a_{r,s}} + \lambda_3 \frac{\partial T_3}{\partial a_{r,s}} \tag{14}$$

We calculate the derivatives of T_1, T_2 and T_3 on $a_{r,s}$:

$$\begin{aligned}
\frac{\partial T_1}{\partial a_{r,s}} &= \frac{\partial}{\partial a_{r,s}} \left[\sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} ||x_i - \mu_p||_{TA} + \sum_{\mu_p \in \mathcal{M}} \sum_{x_i \in \mathcal{C}_p} \sum_{\substack{x_k \in \mathcal{C}_q \\ q \neq p \\ x_i^\phi = x_k^\phi}}^{x_i^t < x_k^t} \left(\beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} (1 - a_{p,q}^2) \right) \right] = \\
&= -2a_{r,s} \sum_{x_i \in \mathcal{C}_r} \sum_{\substack{x_k \in \mathcal{C}_s \\ x_i^\phi = x_k^\phi}}^{x_i^t < x_k^t} \left(\beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} \right) = \\
&= -2a_{r,s}\, pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s), \text{ with the notation } pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s) = \sum_{x_i \in \mathcal{C}_r} \sum_{\substack{x_k \in \mathcal{C}_s \\ x_i^\phi = x_k^\phi}}^{x_i^t < x_k^t} \left(\beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2} \right) .
\end{aligned} \tag{15}$$

$$\frac{\partial T_2}{\partial a_{r,s}} = 2a_{r,s}\, ||\mu_r - \mu_s||_{TA} \ . \tag{16}$$

$$\frac{\partial T_3}{\partial a_{r,s}} = 2a_{r,s}\, inter_\phi^2(\mathcal{C}_p, \mathcal{C}_q) \ . \tag{17}$$

By introducing Equations 15, 16 and 17 into Equations 13 and 14, we obtain:

$$\frac{\partial \mathcal{J}^*}{\partial a_{r,s}} = 2a_{r,s} \left(-\lambda_1\, pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s) + \lambda_2 ||\mu_r - \mu_s||_{TA} + \lambda_3\, inter_\phi^2(\mathcal{C}_p, \mathcal{C}_q) \right) - \lambda \tag{18}$$

We note $K_{r,s} = -\lambda_1\ pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s) + \lambda_2 ||\mu_r - \mu_s||_{TA} + \lambda_3\ inter_\phi^2(\mathcal{C}_p, \mathcal{C}_q)$, we introduce it into Equation 18 and we calculate the fixed point:

$$\frac{\partial \mathcal{J}^*}{\partial a_{r,s}} = 0 \Leftrightarrow 2a_{r,s}K_{r,s} - \lambda = 0 \Leftrightarrow a_{r,s} = \frac{\lambda}{2K_{r,s}} \ . \tag{19}$$

By introducing the constraint in Equation 12 into the Equation 19, we obtain:

$$\sum_{p=1}^{k} \sum_{q=1}^{k} \frac{\lambda}{2K_{p,q}} = 1 \Leftrightarrow \lambda = \frac{1}{\sum_{p=1}^{k} \sum_{q=1}^{k} \frac{1}{2K_{p,q}}} \ . \tag{20}$$

From Equations 19 and 20, we obtain the adjacency matrix update formulas:

$$a_{r,s} = \frac{1}{2K_{r,s}\sum_{p=1}^{k}\sum_{q=1}^{k}\frac{1}{2K_{p,q}}} = \frac{1}{K_{r,s}\sum_{p=1}^{k}\sum_{q=1}^{k}\frac{1}{K_{p,q}}} \ ,$$

$$\text{with } K_{r,s} = -\lambda_1 \ pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s) + \lambda_2||\mu_r - \mu_s||_{TA} + \lambda_3 \ inter_{\phi}^2(\mathcal{C}_p, \mathcal{C}_q) \ ,$$

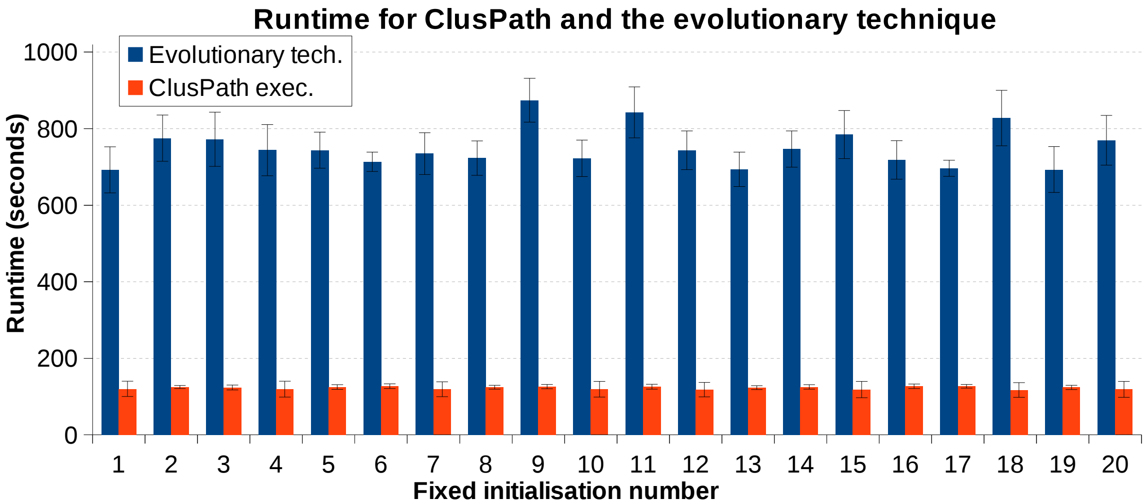
$$\text{and } pen(\mathcal{C}_r \xrightarrow{\phi} \mathcal{C}_s) = \sum_{x_i \in \mathcal{C}_r} \sum_{\substack{x_i^t < x_k^t \\ x_k \in \mathcal{C}_s \\ x_i^\phi = x_k^\phi}} \left(\beta * e^{-\frac{1}{2}\left(\frac{||x_i^t - x_k^t||}{\delta}\right)^2}\right) \ .$$

3 RUNTIME FOR CLUSPATH AND EVOLUTIONARY ALGORITHM

This section reports runtimes for ClusPath and evolutionary algorithm for determining the ClusPath's parameters. All reported runtimes are influenced by the used platform, as well as software and implementation decisions. Consequently, they should be treated comparatively.

All experiments were run on a machine featuring 12 Intel(R) Xeon(R) CPU E5-2430 cores, each running at 2.20GHz. Each core has hyper-threading activated, resulting in 24 logical cores. Each processor disposes of 16MB of cache, while the system has 64GB of RAM space. The system runs Ubuntu Precise 12.04.5, with a 64bit Linux linux kernel version 3.5.0-43.

Both ClusPath and the evolutionary algorithm were programmed in the Matlab/Octave environment and they were run in Octave 3.6.1, using the OpenBLAS 0.1alpha2.2-3 implementation of BLAS (Basic Linear Algebra Subprograms). For the execution of the evolutionary algorithm, OpenBLAS was forced into single threaded execution. The ClusPath algorithm itself is programmed single-threaded (even if it can be easily parallelized, just like most clustering algorithms). Individuals in an evolutionary population are computed in parallel, using the Octave `multicore` package. Therefore, at each generation, the maximum number of ClusPath executions running in parallel at any given time was 24.

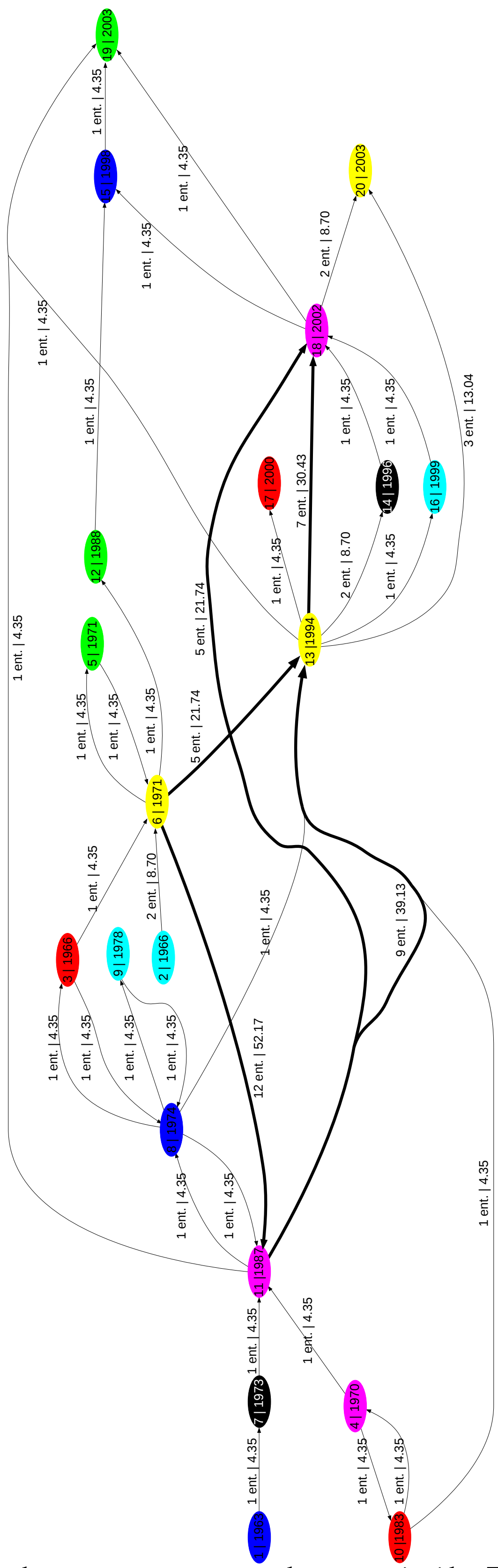


Supplementary Figure 1. Runtimes, in seconds, for ClusPath and the Evolutionary Technique, on `CPDS1`. Each algorithm is run 10 times for each of the 20 initializations presented in Sect. 4.2 of MT.

Supplementary Table 1. Runtimes, in seconds, for "single" ClusPath and the Evolutionary Technique, on `CPDS1`. Each algorithm is run 10 times, for each of the 20 initializations presented in Sect. 4.2 of MT. The table presents the mean of the 10 runs and the standard deviation (in *italic*). The **Ratio** shows how many time slower was the Evolutionary technique compared to "single" ClusPath. The bottom two lines present aggregated measurements over the 20 initializations.

Init. no.	Evolutionary algo.		"single" ClusPath		Ratio
1	692,60	*60,19*	120,19	*19,83*	**5,76**
2	775,30	*60,60*	125,02	*3,92*	**6,20**
3	772,30	*70,70*	123,85	*6,58*	**6,24**
4	744,00	*66,62*	119,51	*20,85*	**6,23**
5	743,90	*47,21*	125,30	*6,23*	**5,94**
6	713,50	*25,08*	127,33	*5,91*	**5,60**
7	735,00	*54,50*	119,04	*19,34*	**6,17**
8	723,20	*44,92*	124,47	*5,17*	**5,81**
9	874,40	*57,35*	126,19	*5,63*	**6,93**
10	722,40	*47,59*	119,06	*20,53*	**6,07**
11	842,70	*66,83*	125,98	*6,34*	**6,69**
12	743,50	*50,49*	118,20	*18,95*	**6,29**
13	693,80	*45,23*	123,50	*5,09*	**5,62**
14	747,00	*47,32*	125,39	*5,71*	**5,96**
15	784,60	*62,89*	118,46	*21,30*	**6,62**
16	718,30	*50,31*	127,02	*5,96*	**5,65**
17	696,60	*20,89*	126,98	*4,85*	**5,49**
18	827,70	*72,36*	117,45	*18,97*	**7,05**
19	693,20	*59,82*	124,17	*5,52*	**5,58**
20	769,70	*65,07*	118,99	*20,76*	**6,47**
Average	**750,69**	***53,80***	**122,81**	***11,37***	6,12
StDev	***51,05***	***13,61***	***3,48***	***7,33***	*0,46*

Figure 2 presents the runtimes of both "single" ClusPath and the evolutionary technique, on `CPDS1`, for each of the 20 initializations presented in Sect. 4.2 of MT. For each initialization, each algorithm is ran 10 times and the means and standard deviation are presented. Furthermore, when running ClusPath alone (*i.e.*, not as an individual in the evolutionary technique), its parameters are chosen randomly from their domain of definition. The purpose of this choice is not to bias the execution of the "single" ClusPath. Fir the first generation of the evolutionary technique, the parameters of the individual runs of ClusPath are chosen similarly, at random from their domain of definition. Therefore, the runtime of ClusPath presented hereafter is comparable to the execution time required for each individual in the evolutionary technique. Table 1 further details these runtimes, presenting also the ratio between the evolutionary technique and "single" ClusPath. More precisely, the evolutionary technique is, in average, 6.12 times slower than ClusPath. That is consistent with our setup: given that each evolutionary population contains 100 executions of ClusPath and 24 executions are done in parallel, the first generations is roughly equivalent to 4 subsequent executions of ClusPath. Given the elitist technique, each later generation requires a total of ClusPath executions less than 24. Given that the entire optimization requires, in average, 3 generations, we obtain a ratio between the evolutionary technique and ClusPath of approximately 6 times.



Supplementary Figure 2. Graph structure constructed *a posteriori* by TDCK-Means, on *Comparative Political Data Set I* with 20 clusters.