

# PREDICCIÓN DEL ERROR DE CONCORDANCIA PLURAL EN CUATRO APRENDIENTES ITALIANOS DE ELE. MATERIAL SUPLEMENTARIO.

## 1. Predicción: atributos basados en grafos y atributos dinámicos

Los siguientes atributos basados en los grafos se calcularon mediante la librería de R “*linkprediction*” (Bojanowski & Chrol 2019). Sean: (i)  $\mathbf{A} = [a_{xy}]$ : la matriz de adyacencia del grafo; (ii)  $n$ : el número de nodos; (iii)  $k_x$ : el grado del nodo  $x$ ; (iv)  $paths_{xy} < l >$ : conjunto de todos los caminos de largo  $l$  del nodo  $x$  al nodo  $y$ ; (v)  $\mathbf{D}$ , la matriz diagonal de los grados; (vi)  $\mathbf{L} = \mathbf{D} - \mathbf{A}$ : el laplaciano del grafo; (vii)  $\mathbf{L}^+$ : la pseudo-inversa de *Moore-Penrose* del laplaciano; (viii)  $\lambda_1$ : el primer autovalor de la matriz  $\mathbf{A}$ ; (ix)  $m$ : el número de enlaces. Se definen entonces las siguientes medidas, que caracterizan los enlaces del grafo (o sea, las concordancias) en términos de la proximidad / similaridad  $S = [s_{xy}]$  entre los nodos conectados por el enlace.

- “preferential attachment” [*pa*]:  $s_{xy} = k_x \times k_y$ ;
- “Kant index” [*ka*]: Conteo de todos los caminos que pasan por un determinado par de nodos.  $S = (I - \beta A)^{-1} - I$ , donde el parámetro  $\beta$  debe satisfacer que:  $\beta < \frac{1}{\lambda_1}$ .
- “L cosine” [*cos\_l*]: Coseno del ángulo entre las columnas de  $\mathbf{L}^+$ .  $s_{xy} = \frac{l_{xy}^+}{\sqrt{l_{xx}^+ l_{yy}^+}}$ .
- “laplaciano inverso” [*l*]:  $S = \mathbf{L}^+$  también proporciona una medida de proximidad entre los nodos.
- “random walk with restart” [*RWR*]: considérese un caminante aleatorio que comienza en el nodo  $x$  y regresa al mismo nodo con probabilidad  $\alpha$ . Sea  $q_x$  la distribución estacionaria de una *cadena de Markov*, es decir:  $q_x = (1 - \alpha)P^T q_x + \alpha e_x$ ; donde:  $e_x$  es un vector con 1 en la posición del nodo  $x$  y  $P$  es la matriz de transición:  $P_{xy} = \frac{1}{k_x}$  si  $A_{xy} = 1$ . La solución para todos los nodos en simultáneo reza:  $q = \alpha(I - (1 - \alpha)P^T)^{-1}$ . Con lo cual se define:  $s_{xy} = q_{xy} + q_{yx}$ .
- “Matrix Forest Index” [*mf*]:  $S = (I + L)^{-1}$ .
- “Average Commute Time” [*act*]: Sea  $r(x, y)$  el número de pasos promedio necesario para que un caminante aleatorio vaya del nodo  $x$  al nodo  $y$ . Para lograr simetría la medida se toma en ambas direcciones de los enlaces:  $n(x, y) = r(x, y) + r(y, x) = 2m(l_{xx}^+ + l_{yy}^+ + 2l_{xy}^+)$ . Entonces:  $s_{xy} = \frac{1}{n(x, y)}$ .
- “Normalized Average Commute Time” [*act\_n*]: Medida anterior pero ahora tomando en cuenta el grado de los nodos.  $s_{xy} = \frac{1}{m(x, y)\pi_y + m(y, x)\pi_x}$ ; donde  $\pi$  es la distribución estacionaria de una *cadena de Markov* que describe un caminante aleatorio en el grafo.

También se incluyó la *frecuencia del enlace* (cuántas veces aparece una determinada concordancia).

En lo que atañe a los atributos dinámicos, se calculó la cantidad de errores acumulados hasta la instancia de concordancia anterior a la instancia  $n$ -ésima. Sobre dicho vector se aplicó una ventana móvil de largo  $w = 45$ <sup>1</sup> y se calcularon las medidas que se describen a continuación.

- Atributos basados en estadística descriptiva<sup>2</sup>. Se calcularon los siguientes estadísticos sobre la distribución de valores de la serie: (a) mediana; (b) media; (c) media podada (5 %); (d) desvío estándar; (f) kurtosis; (g) asimetría; (h) coeficiente de variación; (i) valor máximo; (j) valor mínimo; (k) “*Burtisness*”:  $(\sigma - \mu)/(\sigma + \mu)$ ; (l) “*CV\_diff*”: el desvío estándar de la primera diferencia de la serie:  $\sigma(diff(x))$ ; (m) “*Ben\_dist*”: la media de la porción de la serie que está arriba de la media general relativa a la porción que está debajo de la media general:  $\mu(x_{i>\mu})/\mu(x_{i<\mu})$ ; (n) “*lumpiness*”: varianza de las varianzas; (ñ) “*stability*”: varianza de las medias; “*FE*”: tiempo hasta el valor cuya primera diferencia no es nula; (o) “*crossing\_points*”: número de veces en que la serie cruza su valor medio; (p) “*Max\_level\_shift*”: el cambio en media más grande entre dos ventanas consecutivas.
- Atributos basados en autocorrelación: Se calcula la función de autocorrelación “*ACF*” (detecta correlaciones lineales) y se calcula su valor en el primer *lag* y el tiempo hasta que se produce el primer mínimo y se cruza el primer *cero*.
- “*Hurst*”: índice  $H$  de dependencia a largo plazo de la serie (“memoria”);  $H = 0$  implica incrementos independientes y  $H \in [0.5,1]$  implica “memoria” en la serie.
- “*Motiv*”: Se binariza la serie de manera que los valores mayores a la media son iguales a *uno* y los menores son *cero*. Luego se calcula la entropía de las “palabras” de largo igual a *tres*.
- Atributos basados en *RQA*: Se calculan algunas medidas de *Recurrence Quantication Analysis* (Webber & Marwan 2015), una técnica que permite medir propiedades de atractores en el espacio de fase [phase space] reconstruido. *RQA* mide el grado de nuevas visitas del sistema en estas regiones del espacio de fase y computa ciertas medidas basadas en dichas recurrencias. Específicamente se calculan: (a) recurrencia (“*RR*”): porcentaje de puntos que recurren; (b)

---

<sup>1</sup> Es decir que las primeras 44 observaciones recibieron datos faltantes.

<sup>2</sup> Tanto estos atributos como varios de los que siguen fueron implementaciones hechas en R de los scripts para *Matlab* propuestos en: [www.comp-engine.org](http://www.comp-engine.org). También se usó la librería de R “*tsfutures*”.

determinismo (“DET”): mide la predictibilidad del sistema; (c) Trapping time (“TT”): tiempo promedio en el cual el sistema se encuentra atrapado en un mismo estado de recurrencias; (d) Laminaridad (“LAM”): relacionada con segmentos de tiempo en los cuales no hay intermitencia en el sistema.

- Atributos basados en Entropía de Permutación: se calculó una medida de entropía ( $H$ ) y otra de complejidad ( $C$ ) para la serie. La entropía de permutación fue originariamente propuesta por Bandt & Pompe (2002).
- Atributos basados en “onditas” (*wavelets*): Se aplicó el algoritmo *wavMODWT* (librería de R: *wmtsa*), forma no-decimada de la transformada discreta de *wavelets* (Percival & Walden 2008; Nason 2008). Dados  $n = 2^J$  observaciones ( $J = \log_2(n)$ ), se obtienen como resultado los coeficientes  $s_{jk}$  y  $d_{jk}$  para los niveles  $j = J, J-1, \dots, 1$  y los valores de cada nivel  $k = 1, \dots, 2^{j-1}$ . Los  $s_{jk}$  son los coeficientes de escala (o padre), en cambio los  $d_{jk}$  son los coeficientes de *wavelets madre* (aquí se usa el *Haar*) y se denominan “detalle” [*detail*] porque representan datos de más alta frecuencia. Formalmente:

$$d_{j,k} = \sum_{l=-\infty}^{\infty} g_l y_{2k-l}$$

donde:  $y = (y_1, y_2, \dots, y_n)$  es el vector de datos y  $g_l = \begin{cases} 2^{-1/2} & \text{si } l = 0 \\ -2^{-1/2} & \text{si } l = 1, \\ 0 & \text{si no} \end{cases}$

igualmente para los coeficientes de escala:

$$s_{j,k} = \sum_{l=-\infty}^{\infty} h_l y_{2k-l}, \text{ donde: } h_l = \begin{cases} 2^{-1/2} & \text{si } l = 0 \\ 2^{-1/2} & \text{si } l = 1 \\ 0 & \text{si no} \end{cases}$$

A partir del conjunto de coeficientes para cada nivel, se calculó la energía y la varianza explicada (en escala logarítmica).

Para la frecuencia de los nodos que intervienen en los enlaces de la red se aplicó lo siguiente. Se creó una matriz  $W_{d \times t}$  de filas  $t$ = cantidad de nodos (palabras) de la red y columnas  $d$  = cantidad de concordancias (enlaces de la red). En dicha matriz la celda  $w_{t,d}$  contaba la frecuencia de la palabra  $t$  en la concordancia  $d$ . Se normalizaron las celdas según el esquema:

$$w_{t,d} = tf \times idf = tf_{t,d} \times \log_2 \left( \frac{N}{df_t} + 1 \right)$$

Donde:  $N$  es el número de documentos,  $df_t$  es el número de concordancias que contienen el término  $t$ ,  $tf_{t,d}$  es el número de ocurrencias del término  $t$  en la concordancia  $d$ . Se tomaron las 50 palabras más frecuentes como predictores.

Se llevó a cabo una primera selección de 80 atributos por medio de un *ensemble* de medidas de selección, a saber:

- Área bajo la curva ROC (AUC): Gráfica en la que el eje “y” contiene la proporción de verdaderos positivos y el eje “x”, la proporción de falsos positivos. Si el predictor separara perfectamente las clases habría un corte de probabilidad de asignación de la clase cuyo valor de verdaderos positivos sería 1 y de falsos positivos sería 0; por lo tanto, el área total bajo la curva sería  $AUC = 1$  también. Por otro lado, si un predictor no separa bien las clases, el área se acercaría a  $AUC = 0.50$ . Se debe maximizar.
- Relief Score (Kira & Rendell 1992; cfr. también Kuhn & Johnson 2013: cap. 18): el algoritmo selecciona un punto al azar y luego busca los puntos más cercanos de ambas clases (denominados “error” [*miss*] y “acierto” [*hit*]). Luego se calcula una medida de *score* de la diferencia entre el punto elegido y los errores y aciertos. Esto se repite con  $m$  puntos y el *score* general es una suma de estas diferencias. La intuición radica en que un predictor que separa bien las clases tendrá los “aciertos” cerca y los “errores” lejanos; por eso, *scores* generales altos indicarán predictores influyentes.
- Pesos de Random Forest [RF]: Se calculan los pesos para cada atributo utilizando un modelo de *Random Forest*.
- praznik JMI: La Información mutua indica la reducción en la incertidumbre de una variable si se sabe el valor de otra<sup>3</sup>:  $I(X, Y) = H(X) - H(X | Y)$ . El método selecciona la variable de máxima información mutua entre un predictor  $X$  y la variable respuesta  $Y$ . Luego se añaden uno a uno el resto de los predictores  $X$  tratando de maximizar la función:  $J(X) = \sum_{W \in S} I(X, W; Y)$ , donde  $S$  es el conjunto de las predictoras ya seleccionadas.

## 2. Selección de modelos

Se pueden usar las siguientes medidas para seleccionar modelos (Stroup 2013: 193; Burnham & Anderson 2010: caps. 2 y 4) [ $\theta = \hat{\sigma}, \hat{\beta}$  es el vector de los coeficientes fijos y aleatorios estimados]:

(a) Schwarz:  $BIC = -2L(\theta) + (p \times \log(s))$  [ $s$ = número de grupos; y  $p = p_{\sigma} + (p_{\beta} = \text{rank}[X])$ ]; o sea el número de parámetros fijos más los aleatorios]. Menos es mejor.

---

<sup>3</sup> Recuérdense las definiciones de: (i) entropía (incertidumbre en una variable):  $H(X) = \sum_{x \in X} p(x) \log[p(x)]$ ; (ii) entropía condicional (incertidumbre de una variable sabiendo el valor de otra):  $H(X | Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log[p(x | y)]$ .

(b) Akaike:  $AIC = -2L(\theta) + 2p$ . Menos es mejor.

(c) Akaike corregido:  $AIC_c = -2L(\theta) + 2p(n^*/(n^* - p - 1))$  [  $n^* = N$ , tamaño muestral]; Corrige por muestra pequeña. Menos es mejor. Como heurística, se debería usar cuando<sup>4</sup>:  $\frac{n}{p} < 40$ .

(d) Delta de Akaike:  $\Delta AIC = \Delta = AIC_i - AIC_{min}$ . Indican la distancia del modelo al mejor de todos (el de menor AIC).  $\Delta_i \leq 2$  indica evidencia substancial para el modelo  $i$ .

(e) Pesos de Akaike ( $\omega_i$ ): indican el peso de la evidencia en favor de que modelo sea el mejor de entre todos los modelos candidatos. Es decir, responde a la pregunta: ¿Cómo soportan los datos al modelo  $i$  con respecto al resto de los modelos? Se define como:

$$\omega_{modelo_i} = \frac{\exp(-\Delta_i/2)}{\sum_{i=1}^R \exp(-\Delta_i/2)}$$

donde  $i = 1, \dots, R$  son los modelos considerados; y  $\sum_{i=1}^R \omega_i = 1$ .

(g) Ratio de evidencia (“Evidence Ratio”, [ER]): Ratio entre el peso de Akaike del modelo  $i$ -ésimo y el peso de Akaike del  $j$ -ésimo modelo:  $\frac{w(i)}{w(j)}$ . Muchas veces resulta de interés establecer  $i$  como el índice del mejor modelo:  $\frac{w(1)}{w(j)}$ . Los *ER* son invariantes a los demás modelos, a parte de  $i$  y  $j$ . Responden a la pregunta: ¿Cuántas más veces apoyan los datos al (mejor) modelo  $i$  respecto del modelo  $j$ ?

Una vez ordenados los modelos según alguno de los criterios, se puede reducir dicho conjunto por medio de un “conjunto de confianza” [*confidence set*] para el mejor modelo hallado. Burnham & Anderson (2010: 169) plantean tres alternativas: (i) sumar los pesos de Akaike de los modelos hasta alcanzar  $\geq 0.95$  (recuérdese que los pesos de Akaike suman 1); (ii) tomar los modelos tal que  $\Delta_i \leq 2$ , ya que indican evidencia sustancial para el modelo  $i$ ; (iii) establecer un corte usando ratios de evidencia (poniendo ahora el mejor modelo en el denominador), tal que<sup>5</sup>:  $\frac{w(i)}{w(1)} > \frac{1}{8}$  ( $\Delta_i = 2$ ). Los autores prefieren el tercer criterio debido a su invariancia por adición o borrado de modelos del conjunto de confianza.

Resulta imperativo tener en cuenta la incerteza debida al proceso de selección de modelos. De  $R$  modelos considerados se selecciona el mejor modelo  $i$ . Sin embargo, ¿Si hubieran cambiado los datos, se elegiría igualmente el modelo  $i$  como el mejor o habría variabilidad de entre las muestras de datos en cuanto al modelo elegido?. Una forma de tener en cuenta dicha incerteza es estimar la probabilidad de que un determinado predictor  $x_j$  esté en el mejor modelo si se pudiera recoger una nueva muestra de datos.

---

<sup>4</sup> Las medidas  $AIC$  y  $AIC_c$  convergen para  $n$  grande (manteniendo  $p$  constante). Es decir que cuando dicho ratio es suficientemente grande, tienden a seleccionar el mismo modelo. Entonces, en la práctica conviene usar siempre  $AIC_c$ .

<sup>5</sup> También podrían usarse:  $0.135$  ( $\Delta_i = 4$ );  $0.082$  ( $\Delta_i = 5$ );  $0.05$  ( $\Delta_i = 6$ )

Se trata de una medida de importancia relativa de los predictores. Se lleva a cabo sumando los pesos de Akaike de los modelos en los cuales el predictor  $x_j$  está presente:  $W_+ = \sum w_i I_j(g_i)$ ; donde  $I_j(g_i)$  es la función indicadora que es “1” si  $x_j$  está en el modelo  $g_i$  o cero, si no. Entonces, la importancia relativa es la proporción de modelos en los cuales la predictora está presente.

Si se diera el caso de que, por ejemplo,  $w(i) > 0.9$ , entonces el modelo  $i$  es un claro ganador. En dicho caso es válido hacer inferencia mediante la estimación de los coeficientes  $\beta_i$  y sus errores típicos serán condicionales al modelo seleccionado. Si embargo, muchas veces, especialmente si el conjunto de modelos a considerar es grande, los modelos con  $\Delta_i \leq 2$  poseen pesos de Akaike similares o bien deltas de Akaike cercanos al cero. En este caso,  $\beta_i$  puede diferir en los modelos del conjunto considerado. Una solución es usar la información de todos los modelos involucrados mediante un promedio pesado de los coeficientes. En este caso, los errores típicos de los coeficientes estimados no son condicionales al modelo (ganador) en cuestión sino a todo el conjunto de modelos. Por lo tanto, dichos errores típicos “incondicionales” tienen en cuenta la varianza que proviene del proceso de selección de modelos. Para promediar los coeficientes se utilizó:

$$\bar{\beta}_j = \sum_{i=1}^R w_i I_j(g_i) \hat{\beta}_{j,i} = W_+ \hat{\beta}_{j,i}$$

donde:

$$I_j(g_i) = \begin{cases} 1 & x_j \in g_i \\ 0 & x_j \notin g_i \end{cases}$$

y la suma es sobre todos los modelos del conjunto:  $i = 1, \dots, R$ . En este estimador se usan todos los modelos (“full average”), y cuando la predictora  $x_j$  no estuviera presente en un determinado modelo entonces  $\beta_j = 0$ . Tiene la ventaja de “correr hacia cero” [*Shrinkage*] las estimaciones de parámetros presentes en “modelos malos”. La varianza del estimador resulta:

$$\widehat{var}(\bar{\beta}_j) = \left[ \sum_{i=1}^R w_i \sqrt{\widehat{var}(\hat{\beta}_j | g_i) + (\hat{\beta}_j - \bar{\beta}_j)^2} \right]^2$$

y su error típico:  $\sqrt{\widehat{var}(\bar{\beta}_j)}$ .

### 3. Modelo mixto logístico para cada cluster

Se empleó un modelo logístico mixto condicional para la concordancia  $i$  en el grupo  $j$  (el grupo está definido como *la sesión k* anidada en *el alumno g*). Para el *cluster* 1, los índices son:  $k = 1, \dots, 4, 6, \dots, 12$ ;  $g = 1, 2, 4$ ;  $i = 1, \dots, 631$ . Para el *cluster* 2 son:  $k = 1, \dots, 14$ ;  $g = 1, \dots, 4$ ;  $i = 1, \dots, 1191$ .

Distribuciones:  $y_{ij}|v_{0ij}, w_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij})$ ,  $N = 1$ ;  $v_{0ij} \sim N_q(\mathbf{0}, \mathbf{G}_0)$ ;  $\mathbf{w}_i \sim N_{n_i}(\mathbf{0}, \mathbf{R}_i)$ , donde:  $w_i = w_{j1}, \dots, w_{jn_i}$  ( $n_i = 4, \dots, 81$ ); y con  $q = 1, \dots, 15$  para el cluster 1 y para el 2 es  $q = 1, \dots, 37$ .

Matrices de varianza:  $G_0 = \sigma_{v_0}^2 I$ ; y  $R_i = \sigma_{\omega}^2 I$ .

Función de enlace:  $\text{logit} = g(E[y_{ij}|v_{0ij}, w_{ij}]) = \log \left[ \frac{\pi_{ij}}{1 - \pi_{ij}} \right]$

Función inversa:  $\pi_{ij} = g^{-1}(X\beta + Z_v v + Z_{\omega} w)$

Predictor lineal:  $\eta_{ij} = \varphi_{ij} + v_{0ij} + w_{ij}$ . Donde:  $\varphi_{ij}$  representa los efectos fijos;  $v_{0ij}$  son los efectos aleatorios “entre”;  $w_{ij}$  son los efectos aleatorios “dentro”.

Modelo teórico del Cluster 1:

$$g(E[y_{ij}|v_{0ij}, w_{ij}]) = \log \left[ \frac{\pi_{ij}}{1 - \pi_{ij}} \right] = (\beta_0 + v_{0i}) + w_{ij} + \beta_1 \text{ANIM} + \beta_2 C + \beta_3 \text{ESP1}_3 + \beta_4 \text{ESP1}_4 + \beta_5 \text{ESP1}_5 + \beta_6 \text{ESP1}_6 + \beta_7 \text{EST5} + \beta_8 \text{MOD}_1 + \beta_9 \text{MOD}_2 + \beta_{10} \text{MOD}_3 + \beta_{11} \text{MORF}.f_1 + \beta_{12} \text{MORF}.f_2 + \beta_{13} \text{SKEW} + \beta_{13} \text{EST5} + \beta_{14} \text{ESP2}_2 + \beta_{15} \text{ESP2}_3 + \beta_{16} \text{ESP2}_4 + \beta_{17} \text{EST2} + \beta_{18} \text{LDA} + \beta_{19} \text{STEM}$$

Modelo teórico del Cluster 2:

$$g(E[y_{ij}|v_{0ij}, w_{ij}]) = \log \left[ \frac{\pi_{ij}}{1 - \pi_{ij}} \right] = (\beta_0 + v_{0i}) + w_{ij} + \beta_1 \text{ESP1}_3 + \beta_2 \text{ESP1}_4 + \beta_3 \text{ESP1}_5 + \beta_4 \text{ESP1}_6 + \beta_5 \text{ESP2}_2 + \beta_6 \text{ESP2}_3 + \beta_7 \text{ESP2}_4 + \beta_8 \text{EST1} + \beta_9 \text{FAM}. \text{LEX} + \beta_{10} \text{MOD}_1 + \beta_{11} \text{MOD}_2 + \beta_{12} \text{MOD}_3 + \beta_{13} \text{STEM} + \beta_{14} \text{EST6} + \beta_{15} \text{ES}_1 + \beta_{16} \text{ES}_2 + \beta_{17} \text{LDA} + \beta_{18} \text{EST7} + \beta_{19} \text{EST4} + \beta_{20} \text{MORF}.f_1 + \beta_{21} \text{MORF}.f_2$$

Los efectos aleatorios (condicionales) son: (i)  $v_{0i}$ : la desviación del grupo  $i$  de la ordenada al origen; (ii)  $w_{ij}$ : efectos aleatorios “dentro” de cada grupo, con  $w^T = [w_{j1}, w_{j2}, \dots, w_{jn_i}]$ . Los efectos fijos (marginales) son: (i)  $\beta_0$  es la media basal marginal; (ii)  $\beta_j$  ( $j = 1, 2, \dots$ ) son los efectos de las variables independientes. La matriz de varianza - covarianza de los efectos aleatorios “entre” como aquella de los efectos aleatorios “dentro” es la identidad.

## Referencias

- Bandt, C. & P. Bernd. 2002. “Permutation Entropy: A Natural Complexity Measure for Time Series”, en: *Physical review letters* 88(17). 174102.
- Bojanowski, M. & C. Bartosz. 2019. “Proximity-based Methods for Link Prediction in Graphs with R package ‘linkprediction’”. Disponible en: <http://recon.icm.edu.pl/wp-content/uploads/2019/05/linkprediction.pdf>.

- Burnham, K. P., & D. R. Anderson. 2010. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- Kira, K. & L. A. Rendell. 1992. "The Feature Selection Problem: Traditional Methods and a New Algorithm", en: *Proceedings of the Tenth National Conference on Artificial Intelligence*, AAAI'92, 129-134. San José, California: AAAI Press.
- Kuhn, M. & K. Johnson. 2013. *Applied Predictive Modeling*. Springer.
- Nason, G. P. 2008. *Wavelet methods in statistics with R*. Springer.
- Percival, D. B. & A. T. Walden 2008. *Wavelet methods for time series analysis*. Cambridge University Press.
- Stroup, W. W. 2013. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press: Chapman Hall.
- Webber, C. L. (Jr.) & N. Marwan (2015). *Recurrence Quantification Analysis*. Springer.