

Demografické změny krajů České republiky mezi lety 2006–2011 z pohledu shlukové analýzy¹ Demographic Changes in Regions of the Czech Republic between 2006–2011 as seen by Cluster Analysis

Ondřej Šimpach, Jitka Langhamrová

Abstract: The aim of the paper is to examine the similarity of regions in the Czech Republic according to various indicators from the area of demographic indicators using hierarchical cluster analysis method. Czech Republic has 14 regions in total. Regions are merged into the clusters according to the selected indicators using Euclidean distances. Selected attributes include the number of inhabitants, percentage of population aged 65+ in the total population, live births, deaths total, number of immigrants and number of emigrants (all in 31 Dec). The hierarchical clusterization of regions is calculated for each year based on data of 2006–2011 and next comparison is performed. Ascertained outputs can be used to plan community development and for urban planning such as transport and communications constructions, building of nurseries and basic schools and for decisions about placement of cultural facilities.

Abstrakt: Cílem předkládaného článku je prozkoumat podobnosti krajů České republiky podle různých ukazatelů z oblasti demografie s pomocí metody hierarchického shlukování. Česká republika má celkem 14 krajů. Tyto kraje budou spojeny do několika shluků v závislosti na vybraných indikátorech s využitím Euklidovské vzdálenostní metriky. Zvolené atributy zahrnují počty obyvatel v kraji, procentní zastoupení osob 65+ v populaci, živě narození celkem, zemřelí celkem, počet přistěhovalých a počet vystěhovalých (vše k okamžiku 31. prosince). Hierarchické shlukování krajů je vypočteno pro každý rok z období 2006–2011 a odlišné výsledky jsou spolu vzájemně porovnány.

Key words: Demographic indicators, Ward's method, Euclidean distances, Hierarchical Cluster analysis.

Klíčová slova: Demografické ukazatele, Wardova metoda, Euklidovské vzdálenosti, Shluková analýza.

1. Úvod

Nejenom pro účely územního plánování a rozhodování o investicích ve veřejném sektoru, ale i pro zjednodušování administrativních a ekonomických procesů je výhodné, známe-li podobnost vybraných územních celků navzájem na základě znalostí určitých socio-ekonomických faktorů. O investicích ve veřejném sektoru pojednává např. Nutt, (2006), který byl inspirací pro analýzu politiky soudržnosti určitých územních celků (viz např. Pechrová, Kolářová, 2012). Předkládaná studie čerpá inspiraci zejména od Lv et al., (2011), kteří ve své analýze využili obdobné socio-ekonomické ukazatele pro vytvoření shluků daných územních celků, nicméně jejich analýza byla zaměřena na populace pouze městského typu. Autoři Ozus et al., (2012) využili hierarchického shlukování pro hodnocení efektivnosti výstavby multifunkčních obchodních center na území města. K jejich analýze bylo zapotřebí statistik o vývoji počtů zaměstnaných a nezaměstnaných osob v letech 1970–2000 a dále statistik z cestovního ruchu. Cílem této studie je prozkoumání podobnosti krajů České republiky, podle vybraných demografických ukazatelů (viz Lv et al., 2011) na základě hierarchické shlukové analýzy (Ward, 1963). Zjištěná podobnost může být využita k vysvětlení některých souvislostí (či naopak protikladů), se kterými se můžeme setkat v regionální socio-

¹ Článek byl podpořen z projektu Vysoké školy ekonomické v Praze IGA 6/2013 „Hodnocení výsledků metod shlukové analýzy v ekonomických úlohách“.

hospodářské statistice, či administrativních a rozhodovacích procesech veřejného sektoru, (kterými se mj. zabýval např. Feldstein, 1964).

Česká republika má celkem 14 krajů (jednotek NUTS 3 příslušné klasifikace), jejichž výčet je uveden v Tabulce 1.

Tab. 17: Kraje České republiky (s definovanými zkratkami)

Hl. m. Praha	Hlavní město Praha	KHR	Královéhradecký
STČ	Středočeský	PAR	Pardubický
JlČ	Jihočeský	VYS	Vysočina
PLZ	Plzeňský	JIM	Jihomoravský
KVA	Karlovarský	OLM	Olomoucký
ÚST	Ústecký	ZLN	Zlínský
LIB	Liberecký	MSL	Moravskoslezský

Kraj Hlavní město Praha vychází ve většině publikovaných prací jako odlehle pozorování, v případě shlukování dojde nejspíše k vytvoření jediného a vzdáleného samostatného shluku (viz např. Řezanková et al., 2011 nebo Löster, 2012). Hierarchické shlukování krajů bude vypočteno na základě vybraných údajů (Lv et al., 2011 nebo Arnio, Baumer, 2012) z let 2006, 2007, ... a 2011, přičemž data byla pořízena z databáze Českého statistického úřadu (ČSÚ) a databáze Ministerstva zemědělství (MZe). Databáze MZe s podrobností na obce byla v minulosti využita např. i k analýze vybraných okresů (Šimpach, 2013). Vývoj těchto shluků bude tedy možné porovnat v šestiletém časovém horizontu.

2. Metodika a data

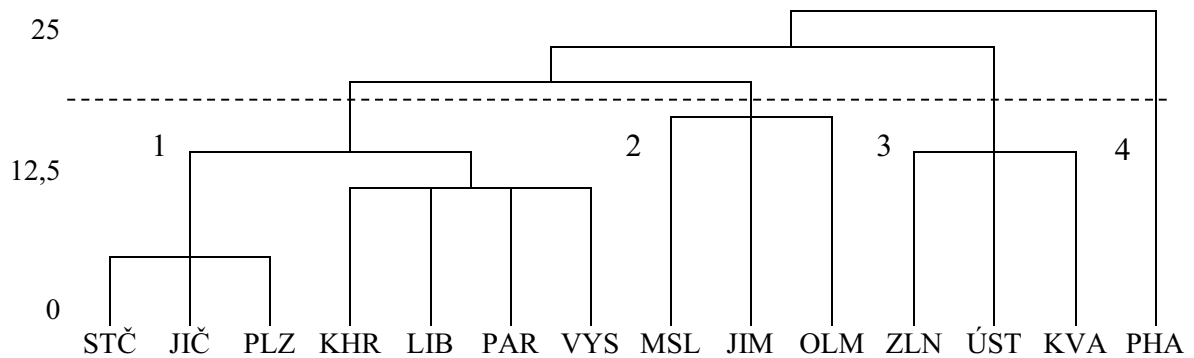
Vzdálenosti mezi jednotlivými kraji České republiky jsou vypočteny s využitím Euklidovské metriky (čtvercových vzdáleností). Poté jsou na základě známých matic vzdáleností rozděleny do 4–5 shluků (s ohledem na vhodnost zařazení do příslušného shluku (viz např. Löster, 2012)) a pochopitelně v závislosti na vybraných demografických ukazatelích, s využitím hierarchického shlukování a Wardovy metody (viz např. Danielson, 1980 nebo Bavaud, 2010). Počty shluků vychází ve většině případů 4 a vyplývají z dendrogramů, jejichž řez byl proveden vždy na stejné vzdálenosti, aby byly jednotlivé výsledky mezi sebou srovnatelné a dále pak z doporučení udané CHF indexem, zvaným též pseudo F index (viz Calinski, Habarasz, 1974 a dále aplikace Löstera, 2011), založeném na podílu průměrné mezishlukové a průměrné vnitroshlukové variability. Ze zmíněných datových matic, pořízených z databází ČSÚ a MZe, byly vybrány na základě zmiňovaných literárních zdrojů statistiky o

- počtu obyvatelích v daném kraji,
- podílu osob 65+ v populaci,
- počtu živě narozených celkem,
- počtu zemřelých celkem,
- počtu přistěhovalých a
- počtu vystěhovalých,

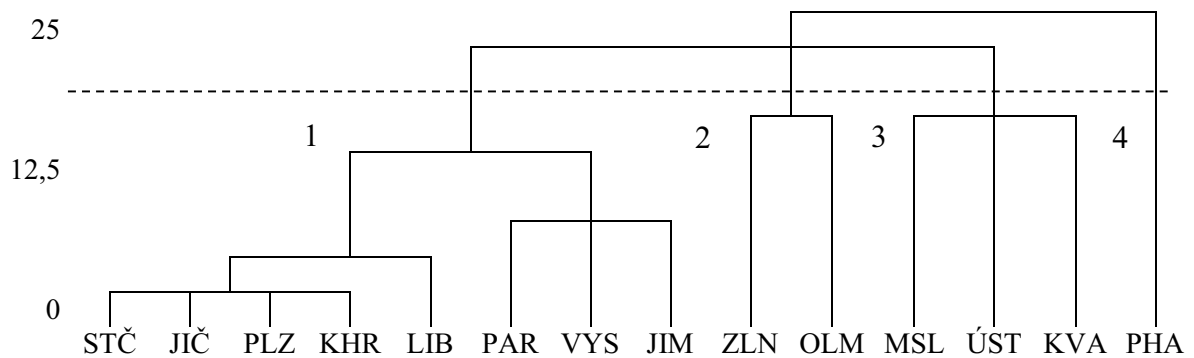
vše aktuální k 31. prosinci roků 2006, 2007, ... a 2011. Výpočty vzdálenostních matic byly prováděny v systému IBM SPSS Statistics, na základě nichž byly konstruovány dále prezentované dendrogramy.

3. Výsledky

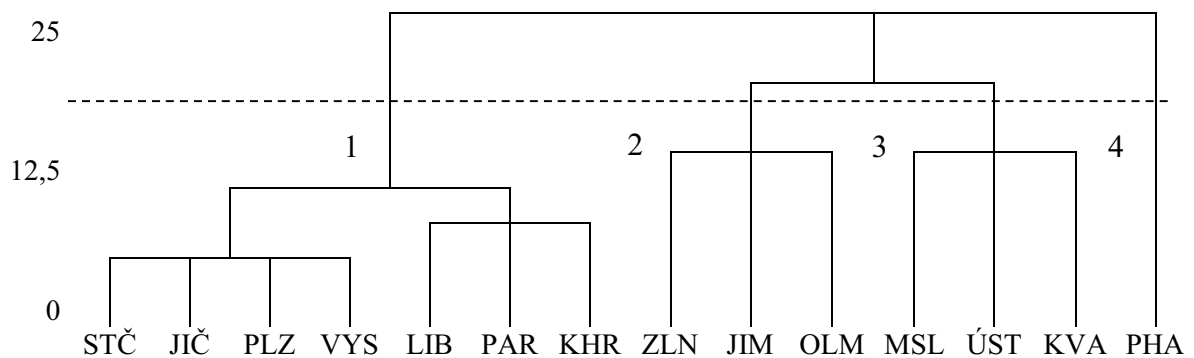
S využitím z -transformace, hierarchického shlukování založeném na Wardově metodě (Ward, 1963) a čtvercových Euklidovských vzdálenostních metrikách byly vypočteny shluky pro kraje České republiky v letech 2006, 2007, ... a 2011. Řezy dendrogramy byly provedeny na vzdálenosti 19 jednotek, čímž došlo ve většině případů k vytvoření 4 shluků, v jednom 5. Dendrogramy jsou postupně zobrazovány v obrázcích 1–6.



Obr. 1: Dendrogram pro kraje ČR v roce 2006. (zdroj: vlastní výpočet a konstrukce)



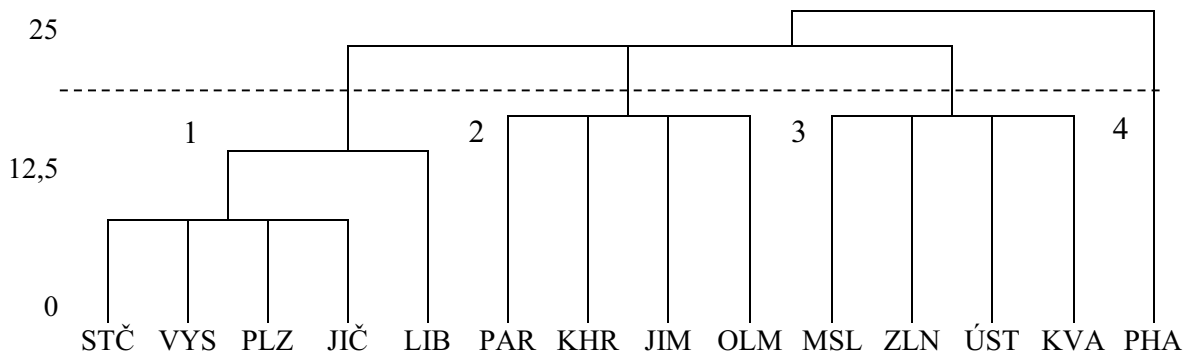
Obr. 2: Dendrogram pro kraje ČR v roce 2007. (zdroj: vlastní výpočet a konstrukce)



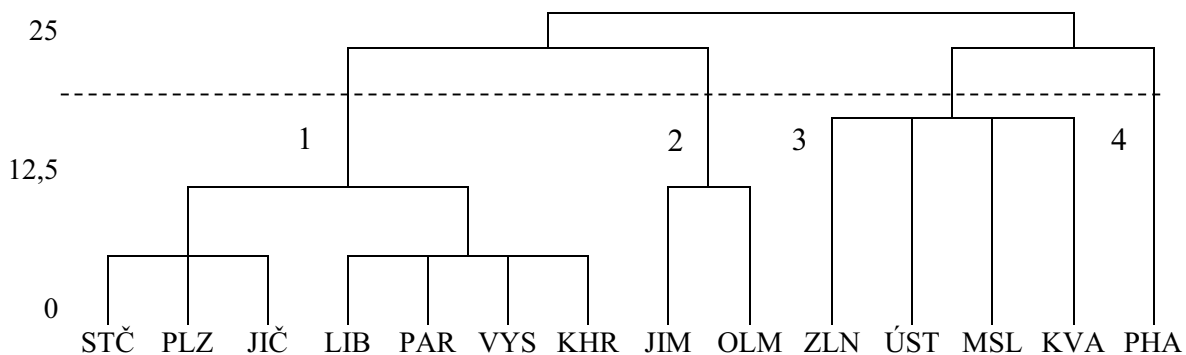
Obr. 3: Dendrogram pro kraje ČR v roce 2008. (zdroj: vlastní výpočet a konstrukce)

Při pohledu na situaci v roce 2006 vidíme pod označením „1“ velký shluk sedmi pouze českých krajů. Ty jsou podobné zejména vyšším podílem seniorů nad 65 let v populaci a nižšími počty živě narozených dětí. Také se jedná o kraje migračně atraktivní a převažuje u

nich kladné migrační saldo. Kraj Karlovarský a Ústecký je zahrnut ve shluku číslo „3“ spolu s jedním moravským krajem – Zlínským. Tyto kraje spolu souvisí zejména vyšší mírou emigrace než ostatní kraje. Kraj Hlavní město Praha tvoří jeden samostatný shluk proto, že je ve většině srovnávaných statistik výrazně odlišný od ostatních krajů. Ve shluku číslo „2“ se nachází kraje pouze moravské – Moravskoslezský, Jihomoravský a Olomoucký. V roce 2007 je zajímavé pozorovat, že do velkého shluku s označením „1“ vstupuje Jihomoravský kraj. Statistiky, které vstoupily do analýzy, byly v roce 2007 v Jihomoravském kraji obdobné, jako u zmínovaných českých krajů. Shluk číslo „2“ je tedy zmenšen o jednoho člena, shluk číslo „3“ zůstal nezměněn, pouze se mírně změnila hodnoty v matici vzdáleností. Kraje Moravskoslezský, Ústecký a Karlovarský mají každoročně mnohem vyšší míry emigrace než všechny ostatní kraje, je to způsobeno zejména horšími pracovními podmínkami v regionech a méně rozvinutou infrastrukturou. V roce 2008 se některé kraje přeuspořádaly v rámci velkého shluku „1“, Jihomoravský kraj se přesunul od českých krajů k moravským do shluku „2“ (ke Zlínskému a Olomouckému). Shluk číslo „3“ zůstal v roce 2008 od předchozího roku nezměněn. Kraj Hlavní město Praha je nejvíce odlišný od ostatních krajů zejména vyšším podílem seniorů nad 65 let v populaci, (jde o regresivní typ populace), a dále vyšší mírou imigrace.

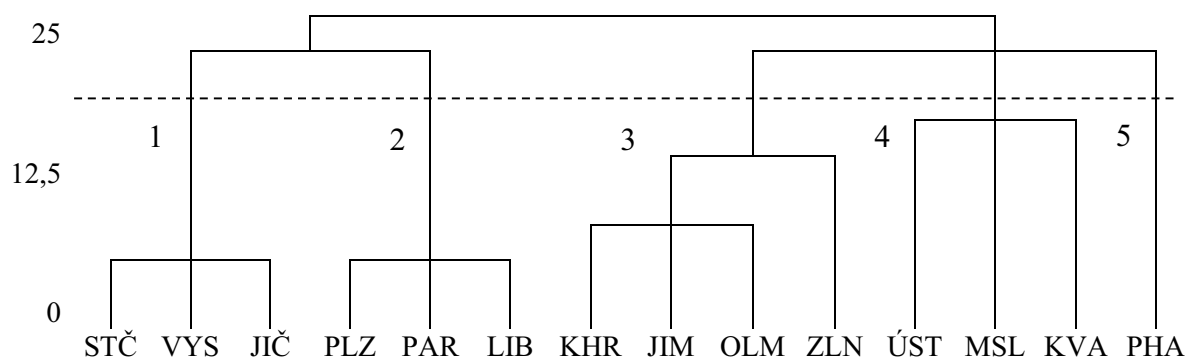


Obr. 4: Dendrogram pro kraje ČR v roce 2009. (zdroj: vlastní výpočet a konstrukce)



Obr. 5: Dendrogram pro kraje ČR v roce 2010. (zdroj: vlastní výpočet a konstrukce)

Rok 2009 byl ve znamení přeuspořádání členů mezi jednotlivými shluky. Uspořádání na obrázku 4 je mnohem rovnoměrnější a shluky jsou podobně velké. Byl to rok silného ekonomického poklesu, který rozhodně měl vliv i na zmíněné ukazatele z oblasti demografie, které rozhodovaly o takovémto výsledku. Nejvíce rovnoměrné byly v tomto roce statistiky imigrace a emigrace jednotlivých krajů. Počty zemřelých se dlouhodobě mění jen nepatrně, ale počty živě narozených na zmíněné události reagovaly.



Obr. 6: Dendrogram pro kraje ČR v roce 2011. (zdroj: vlastní výpočet a konstrukce)

Srovnáme-li rok 2009 s rokem 2010, zjistíme, že rok 2010 byl z pohledu výsledků mnohem více variabilní. Do soustavy se navrátil velký shluk číslo „1“, opět tvořený pouze českými kraji. Jihomoravský a Olomoucký kraj (oba z Moravy) tvoří shluk číslo „2“, ve shluku číslo „3“ zůstaly stejné kraje jako v roce 2009. U Zlínského, Ústeckého, Moravskoslezského a Karlovarského kraje pravděpodobně nedošlo k žádným významnějším změnám z pohledu sledovaných ukazatelů. Hlavní město Praha si drží svou suverenitu vždy ve čtvrtém samostatném shluku, jak v roce 2009, tak i ve zbývajících letech. Poslední dendrogram na obrázku 6 je jediným, kde došlo k rozdělení krajů České republiky do pěti shluků. Původní velký shluk se rozpadl na dva menší. K Jihomoravskému a Olomouckému kraji, které byly v roce 2010 ve shluku číslo „2“ (v roce 2011 už číslo „3“), přibyl Královéhradecký a Zlínský kraj. Ve shluku dříve označeném „3“ (nyní „4“), bývaly a i se nyní nachází kraje Ústecký, Moravskoslezský a Karlovarský. Tyto kraje v roce 2011 opět zaznamenávaly nejvyšší záporná migrační salda – mechanický úbytek obyvatelstva.

4. Diskuse a závěr

Výsledky této studie jsou do velké míry ovlivněny použitou metodikou. Při využití jiné než Wardovy metody bychom dostali shluky jiné (viz např. Löster, 2011). Obdobně pak nenormalizovaná vstupní data mění výsledky zásadní měrou. V našem případě však normalizace byla zapotřebí, neboť vstupní data nebyla ve srovnatelných jednotkách.

Demografické ukazatele, které spolu nejvíce souvisí a spojují jednotlivé kraje, jsou především podíl obyvatel 65letých a starších v celkové populaci a dále pak statistiky migrace. Kraje, které mají již dlouhodobě regresivnější populační strukturu, bývají spojovány do stejných shluků a kraje, které dlouhodobě působí jako migračně neatraktivní pak také. Počty obyvatel a počty zemřelých celkem se v dlouhém období mění jen velmi nepatrně, proto přesouvání krajů mezi jednotlivými shluky nehrozí ze strany těchto statistik. Studie by v budoucnu mohla být rozšířena o další socio-ekonomické ukazatele či národohospodářské indikátory. Např. studie Löstera a Langhamrové (2011) poskytuje mnohé informace o vývoji nezaměstnanosti, a právě míra nezaměstnanosti a další statistiky z trhu práce by byly adekvátním doplňkem pro další kalkulace.

Literatura

ARNIO, Ashley N., BAUMER, Eric P. (2012). Demography, foreclosure, and crime: Assessing spatial heterogeneity in contemporary models of neighborhood crime rates, *Demographic Research*, Vol. 26, (May 2012), p. 449-486.

BAVAUD, F. (2010). Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs, *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, Volume 6321, 2010, pp 103-118.

- CALINSKI, T., HARABASZ, J. (1974). A Dendrite Method for Cluster Analysis, *Communications in Statistics*, No. 3, 1974, pp. 1-27.
- DANIELSON, Per-Erik (1980). Euclidean distance mapping, *Computer Graphics and Image Processing*, Volume 14, Issue 3, November 1980, Pages 227–248.
- FELDSTEIN, Martin S. (1964). Net Social Benefit Calculation and the Public Investment Decision, *Oxford Economic Papers New Series*, Vol. 16, No. 1 (Mar., 1964), pp. 114-131.
- LÖSTER, T. (2011). *Hodnocení výsledků metod shlukové analýzy*. (Doktorská disertační práce). Praha : FIS VŠE v Praze, 2011, 137 s.
- LÖSTER, T. (2012). Kritéria pro hodnocení výsledků shlukování se známým zařazením do skupin založená na konfuzní matici. *Forum Statisticum Slovacum [online]*, 2012, roč. 8, č. 7, s. 85–89. ISSN 1336-7420. URL: <http://ssds.sk/casopis/archiv/2012/fss0712.pdf>.
- LÖSTER, T., LANGHAMROVÁ, J. (2011). Analysis of Long-Term Unemployment in the Czech Republic. Praha 22.12.2011–23.12.2011. In: LÖSTER, Tomáš, PAVELKA, Tomáš (ed.). *International Days of Statistics and Economics*. Slaný : Melandrium, 2011, s. 228–234. ISBN 978-80-86175-77-5.
- LV, J., LIU, QM., REN, YJ., GONG, T., WANG, SF. and LI, LM. (2011). Socio-demographic association of multiple modifiable lifestyle risk factors and their clustering in a representative urban population of adults: a cross-sectional study in Hangzhou, China, *International Journal of Behavioral Nutrition and Physical Activity*, 2011, Vol. 8 : 40.
- NUTT, Paul, C. (2006). Comparing Public and Private Sector Decision-Making Practices, *Journal of Public Administration Research and Theory*, (April 2006) 16 (2): 289-318.
- OZUS, E., AKIN, D., and ÇIFTÇI, M. (2012). Hierarchical Cluster Analysis of Multicenter Development and Travel Patterns in Istanbul, *Journal of Urban Planning and Development*, Volume 138, Issue 4 (December 2012), 303–318.
- PECHROVÁ, M., KOLÁŘOVÁ, A. (2012). Does the cohesion policy mitigate the disparities among the regions in the Czech Republic?. Karviná 09.11.2012. In: *Mezinárodní vědecká konference doktorandů a mladých vědeckých pracovníků [CD-ROM]*. Opava : Slezská univerzita, 2012, p. 224–234. ISBN 978-80-7248-800-1.
- ŘEZANKOVÁ, H., LÖSTER, T., HÚSEK, D. (2011). Evaluation of Categorical Data Clustering. Fribourg 26.01.2011–28.01.2011. In: *Advances in Intelligent Web Mastering – 3*. Berlin : Springer Verlag, 2011, s. 173–182. ISBN 978-3-642-18028-6. ISSN 1867-5662.
- ŠIMPACH, O. (2013). Application of Cluster Analysis on the Demographic Development of Municipalities in the Districts of Liberecký Region. Prague 19.09.2013 – 21.09.2013. In: *International Days of Statistics and Economics at VŠE, Prague*. Prague : VŠE, 2013, s. 1390–1399. ISBN 978-80-86175-87-4.
- WARD, J. H., Jr. (1963). Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58, 236–244.

Adresa autorů:

Ondřej Šimpach, Ing.
KDEM FIS VŠE v Praze
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
ondrej.simpach@vse.cz

Jitka Langhamrová, doc., Ing., CSc.
KDEM FIS VŠE v Praze
Nám. W. Churchilla 4, 130 67 Praha 3
Česká republika
langhamj@vse.cz