Geographia Technica, Vol. 16, Issue 2, 2021, pp 183 to 192

BUILDING AND EXPLORING NETWORK DATA MODEL FOR A SEASON LEVEL CLIMATE CHANGE STUDY FOR FIVE LARGE CITIES IN HUNGARY

Zsolt MAGYARI-SÁSKA[1]

DOI : 10.21163/GT_2021.162.15

ABSTRACT:
Since a few years or decades, climate change has an impact felt in a very visible way in everyday life. People are increasingly confronted with its negative effects. Naturally, scientific research on the subject is multiplying. The present exploratory study attempts to present a network modelling approach on studying climate change. Networks are increasingly present in different areas of life, but have not played a significant role in climate research. This publication attempts to assess climate changes at five municipalities in Hungary by developing and analyzing three network data models. The developed different data models provide an opportunity to approach climate change from different perspectives, as the change itself is multifaceted. Data analyses are based solely on the structural indicators of the constructed networks, the measured weather characteristics only contributing to the construction of the data model. The obtained results for each location and season are complex, but interpreting them together helps to see the variations and their different nature.

Key-words*: Graph model, Data mining, Big data, Complex network, Similarity measure***,** *R-CRAN.*

1. INTRODUCTION AND AIMS

Climate change is now a phenomenon that has largely convinced even the most skeptical. Today, reporting on extreme natural phenomena is no longer just a sensational news, it has become part of everyday life. Ordinary people are also facing the consequences of climate change, as it affects many aspects of their lives, whether it is related to agriculture, tourism or even their everyday security. This has led the scientific community to intensify its efforts to study climate change. Good research requires good data models and the right methods to go with them.

The idea that other research methods could be used to study the climate has been constantly developing. Since the weather that generates the climate implies a succession of data, and these large or small periods, form a temporal chain, logically raise the question of whether, Moran's First Law (Moran, 1950), issuing the spatial similarity could be applied for time (events closer in time are more similar than events further apart in time). Of course, that is not necessarily true, as aggregated data of weather for a later day, week or month may be more similar to a past period with similar duration than weather of a day, week or month immediately before it.

However, the existence of the link, the connection, the correlation between periods that may be distant in time is a real situation, and the study of it may prove useful as it indicates the dynamics of weather change. The study of it need a proper data model. One of the most high-profile scientific innovations of recent years is the use of networks (Barabási, 2013), which has very diverse application areas such as community detection (Li and Maini, 2005; Magyari-Sáska, 2019), medicine and life sciences (Light et al., 2005; Hopkins, 2007), economy analysis (Borgs et al, 2007; Emmert-Streib et al., 2018), security (Arquilla and Ronfeldt, 2001). Missing from all these broad applications is climate assessment.

[1] *Babeș-Bolyai University, Faculty of Geography, Gheorgheni University Extension, 535000 Gheorgheni, zsolt.magyari@ubbcluj.ro*

The aim of this research is to develop a new network-based data model and research methodology for climate research and to apply it in Hungary. At urban level there are just a few climatic studies for Hungary or for other countries (Probáld, 2014; Stone, 2012), but climate change at city level is an important as they are both generating factors and victims of it (Bulkeley, 2012; Hunt and Watkins, 2011). Since in 2021 daily meteorological data series for the period 1901-2020 for five municipalities of Hungary (Budapest, Debrecen, Pécs, Szeged and Szombathely) are available, which is suitable for the climate studies to be carried out, both in terms of time period and level of detail.

The data aggregation period in this study will be at the level of seasons, as several studies have examined the change and shift of seasons (Thompson, 2009; Magyari-Sáska and Dombay, 2020). A network model will be developed in which nodes represent the seasons, while the links between them will be established according to which new node (period) is (are) most similar to the previous period(s). The proper evaluation of this data model allows the assessment of climate change in terms of weather variability.

2. WEATHER DATA AND THE NETWORK MODELS

A set of daily meteorological data between 1901 and 2020 was available from the official site of the National Meteorological Service of Hungary. The dataset includes records on the average, minimum and maximum temperature as well as the precipitation quantity for every day. The original dataset was homogenized as for every meteorological station several relocations and measurement instruments changes has been done over the years. The location of the five cities covers the different regions of Hungary.

The network model used in this study is based on the similarity of the weather parameters between the time periods. In case of this research the nodes in the network are seasons, while the links are placed between them according to the best similarity. To obtain a similarity value based on which the edges of the network can be placed, at first we had to aggregate the weather values. This process has meant an upscale of the original data. The following 32 aggregated values were calculated for every season: average, minimum, maximum, standard deviation, maximum deviation and average of the daily variation of daily mean, minimum and maximum temperatures, maximum of the daily maximum temperature variation, maximum variation over the whole time interval based on daily maximum and minimum, total amount of precipitation, total amount of precipitation from fog, total liquid precipitation, total solid precipitation, number of days without precipitation, number of days with hail, number of days with thunderstorms, number of days with snow showers, number of days with hail and thunderstorms and number of thundery days.

In order to build up network models, a similarity index should be defined to express the combined similarity of different weather characteristics of two aggregation periods. Since the values of each weather characteristic are presented on different scales and orders of magnitude, it was necessary to normalize these values beforehand. A similarity index can be constructed by trying to condense (e.g. using a linear combination) the standardized values of the weather characteristics into a single value and then designating the difference between these values as the similarity index. In this case, the question that rises at what extent the different characteristics are involved in the formation of the weather, what their information content is, whether they need to be weighted and, if so, what these weights should be. Applying this direction would have been very subjective.

The similarity index can be created without combining the weather characteristics into one value. There are several methods to determine the similarity between two vectors, each consisting of several values (in our case weather characteristics). One of the best known is correlation, but there are several others as Jaccard index, multidimensional Euclidean, Minkowski or even Hamming distance, or cosine similarity (Tan et al., 2005). I have used the latter, since this indicator is often used in data mining (Han et al., 2012) and shows how the values of two vectors have the same orientation. By definition, the cosine similarity measure is the cosine of the difference between the outcome vectors defined in a multidimensional (the number of dimensions is equal to the number of features in the vectors) space (eq. 1).

$$sim(v,w) = \frac{v \cdot w}{\|v\| \times \|w\|} = \frac{\sum_{i=1}^{n} v_i \times w_i}{\sqrt{\sum_{i=1}^{n} v_i^2} \times \sqrt{\sum_{i=1}^{n} w_i^2}} \quad [1]$$

After several attempts three network models were adopted to carry out the data analysis. Even if the construction for all of them is based on the similarity index, each tries to catch the reality from a different point of view. The first network model is based on preferential attachment between nodes considering their show up in time. Every new node will be attached to that one in the existing model to which has a highest similarity value. The result will be a tree. The second model disregards the time in the construction process, as every node will be attached the which has the highest similarity value not considering if that node represents a time period before or after the node on process. The result in this case is a disconnected graph. The third network model starts from the complete network (all nodes are connected to each), but at final will remain just those links which are part of the shortest routes between each node pair. The result will be a connected graph. Each of the selected network model has a different view on similarity between the nodes. By respecting the time flow, the first model is more suitable to catch the changes. The second model concentrates on the highest similarity and by this can emphasized the compactness or spread of the climate. The third model wants adds the possibility for a node to be connected to multiple others in order to get the highest similarity to all other nodes. This model also makes possible to study the compactness of the climate but unlike model two it can take the global similarity between node pairs.

The data aggregation process and the network model construction for each of the five municipalities was carried out based on R functions (Ihaka and Gentlman, 1996) as individual and manual data processing was out of question due to the high amount of data. Beside the base R system the igraph package (Csárdi and Nepus, 2006) was used to perform all the network related operations such as network creation, adding edges, determining shortest paths, weighting the graph etc. For network visualization the Cytoscape software were used (Shannon et al., 2003) having some basic network analysis possibilities as well.

3. RESULTS AND DISCUSSION

Using the first data model, which is based on preferential linkage between aggregation periods using highest similarity, all seasons can be identified very clearly for all five municipalities (**Fig. 1**). Using the radial representation in Cytoscape, it is easy to interpret in climate changing terms the longer arms extending outwards or the fan-like shapes spreading out from the center, even if the automatic placement algorithm in some cases does not place them side by side.

The longer the elongations are, the more they represent a new changing direction of the season, while the fan shapes represent a higher stability of the season. On this basis, while Budapest shows a roughly equal variation between seasons, Debrecen shows a pronounced change in the summer compared to the other seasons. As the number of nodes is also significant on the stretching arm, the summer change at Debrecen is a multi-year change, unlike in cases of autumn at Pécs, the spring at Szeged or the winter at Szombathely. In these locations there are some outliers, but they contain contains only a few nodes.

However, some nodes in the network are displaced, being present in the branch of another season category (**Fig. 2**). If we look at these situations summarized in **Table 1**, the cases when a node is not in his seasonal branch, we can find the following highlighted situations:

- winter in spring branch; this situation occurs in all cities except Debrecen and in 44% of the cases in the last 30 years. This type of divergence is particularly marked in the case of Budapest. All these situations indicate mild winters.

- winter in autumn branch; is the case of Budapest alone and occurs only in three cases, in the past, more than 30 years ago, but also indicates mild winters.

- spring in summer branch; we find this situation in case of Budapest and Pécs, indicating warm springs. 71% of these situations have occurred in the last 30 years.

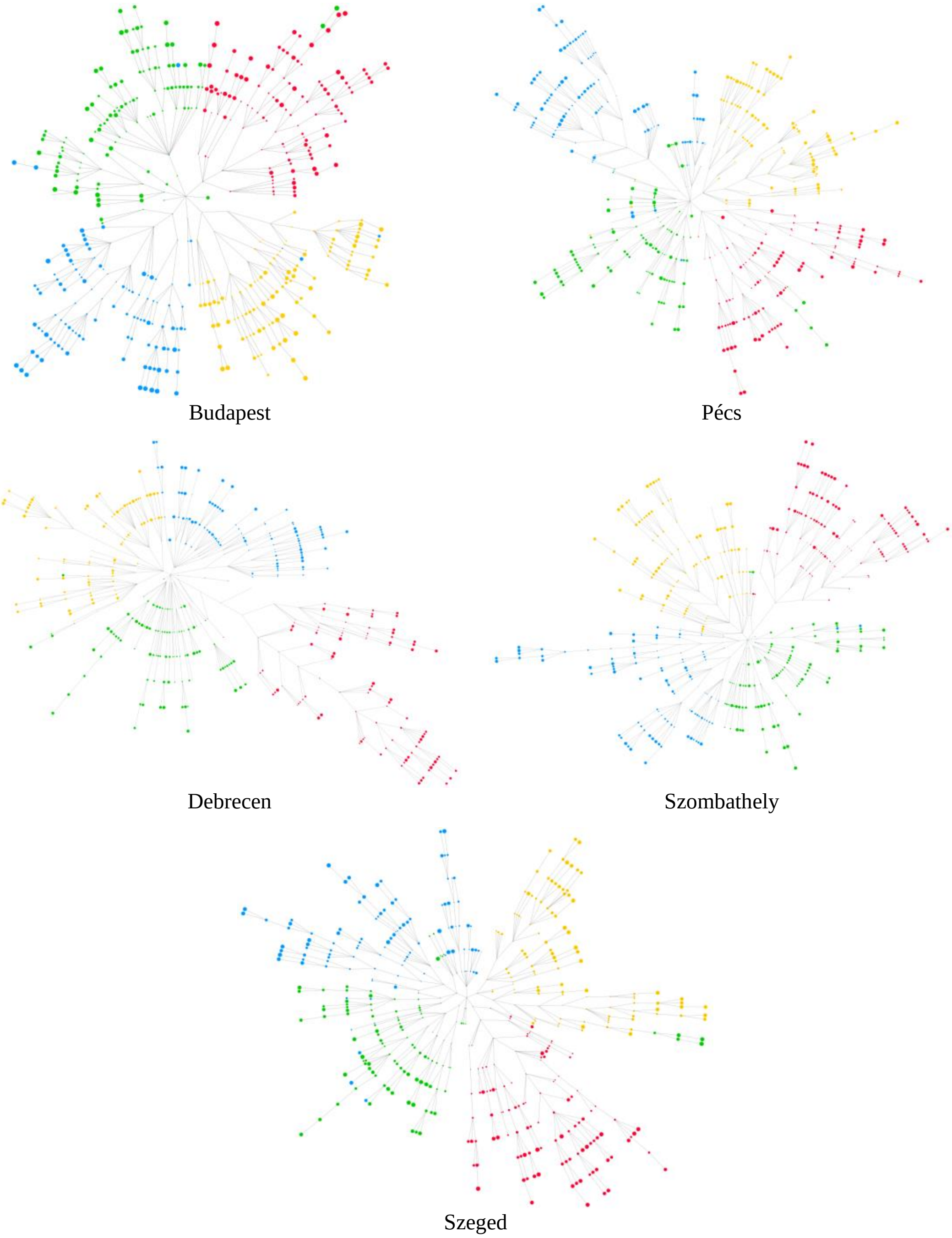


Fig. 1. – Network data models based on preferential connectivity with radial representation
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

It is noticeable that winter branches do not contain other seasons, winter does not "attract" other seasons, but rather in some cases dissolves into spring or autumn.

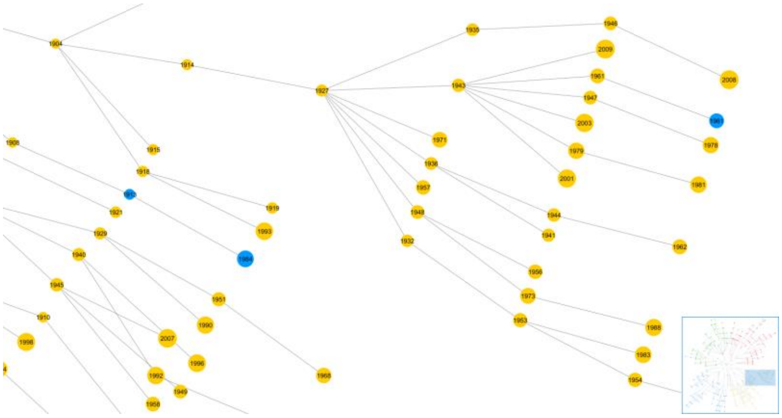


Fig. 2. – Different seasons in a branch (case of Budapest - three winter seasons in autumn branch)

Table 1.

Displaced seasons.

<table>
<tr><th>CITY</th><th>TYPE OF DISPLACEMENTS</th><th>YEARS OF DISPLACEMENTS</th></tr>
<tr><td rowspan="3">Budapest</td><td>winter in spring branch</td><td>1914 | 2003 | 2006 | 2013</td></tr>
<tr><td>spring in summer branch</td><td>2014 | 2019</td></tr>
<tr><td>winter in autumn branch</td><td>1913 | 1961 | 1984</td></tr>
<tr><td>Debrecen</td><td>spring in autumn branch</td><td>1931 | 1940 | 1994 | 1998 | 2012</td></tr>
<tr><td rowspan="3">Pécs</td><td>spring in summer branch</td><td>1934 | 1990 | 1994 | 1998 | 2012</td></tr>
<tr><td>winter in spring branch</td><td>1965 | 1988 | 1997 | 2018 | 2020</td></tr>
<tr><td>autumn in spring branch</td><td>1930 | 1947 | 1964</td></tr>
<tr><td rowspan="2">Szeged</td><td>spring in autumn branch</td><td>1967 | 1989 | 2007 | 2014</td></tr>
<tr><td>winter in spring branch</td><td>1924 | 1926 | 1947 | 1954 | 1988 | 1989 | 2008</td></tr>
<tr><td rowspan="2">Szombathely</td><td>winter in spring branch</td><td>1989 | 1996</td></tr>
<tr><td>spring in autumn branch</td><td>1922 | 1937 | 1945 | 1985</td></tr>
</table>

If we look at the evolution of the number of displacements over 30-year periods (**Fig. 3**), we can see a sharp increase. While in the first 30 years (1901-1930) there were only 6 such situations, in the following period there were 8, then 11, and in the last 30 years 17. For Szombathely and Debrecen, the total number of displacements is lower, but the percentage of displacements for the last 30 years is comparable with registered at other locations. For Szombathely, only 17% of the displacement occur in the last 30 years, while for Debrecen the percentage is 60%. Considering the total number of displacements, Budapest is in the middle of range, but proportionally it has more displacements for the last 30 years than Pécs or Szeged, which have higher total displacements (**Fig. 4**).

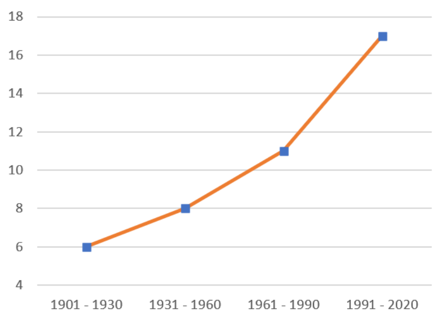


Fig. 3. – Evolution of displaced seasons number

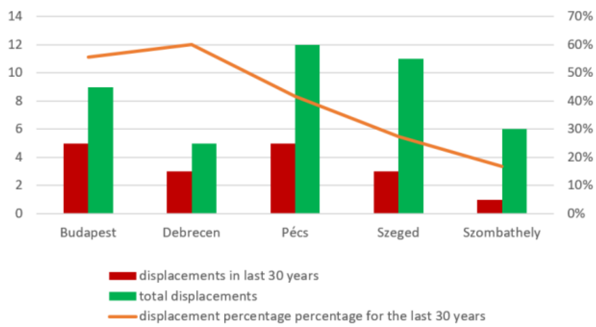


Fig. 4. – Displacement percentage for the last 30 years

The visual content of the second data model is different from the previous one (**Fig. 5**). In this case the number of components and the number of nodes in these components are the most relevant. The more components has a network, the more heterogeneous was the weather in the given municipality. However, not only the number of components is important, but also the number of nodes. The larger components are placed at the beginning of each diagram. Components with at least 10% of the total number of nodes has been named as large components, which characterize the most similar weather conditions.

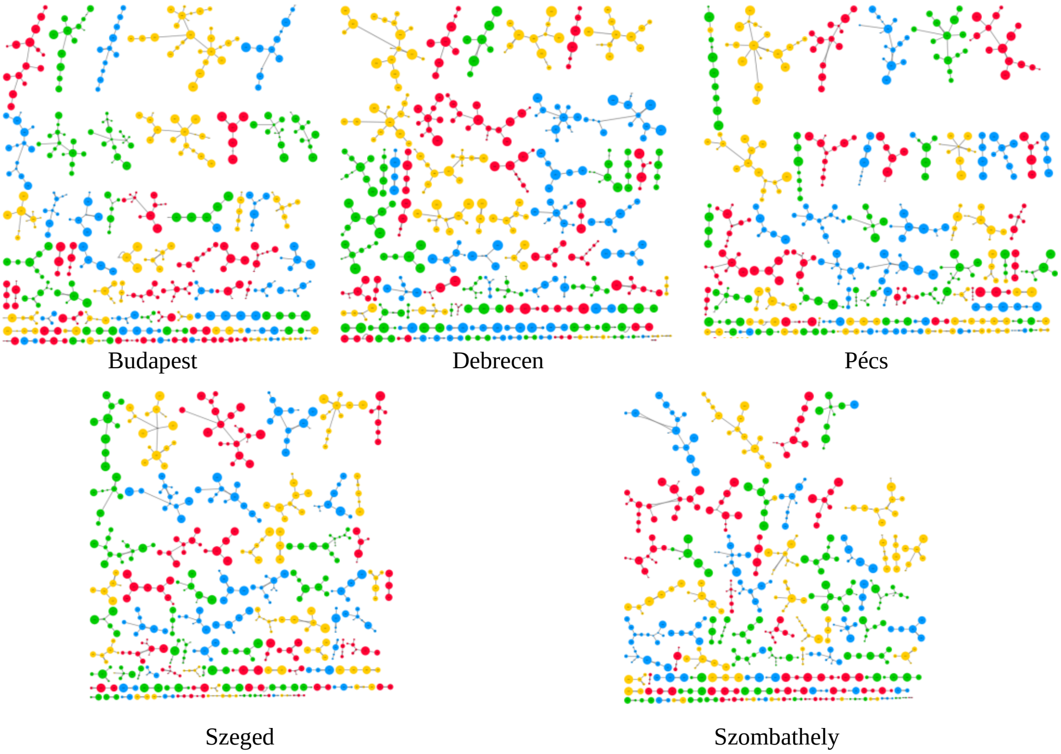
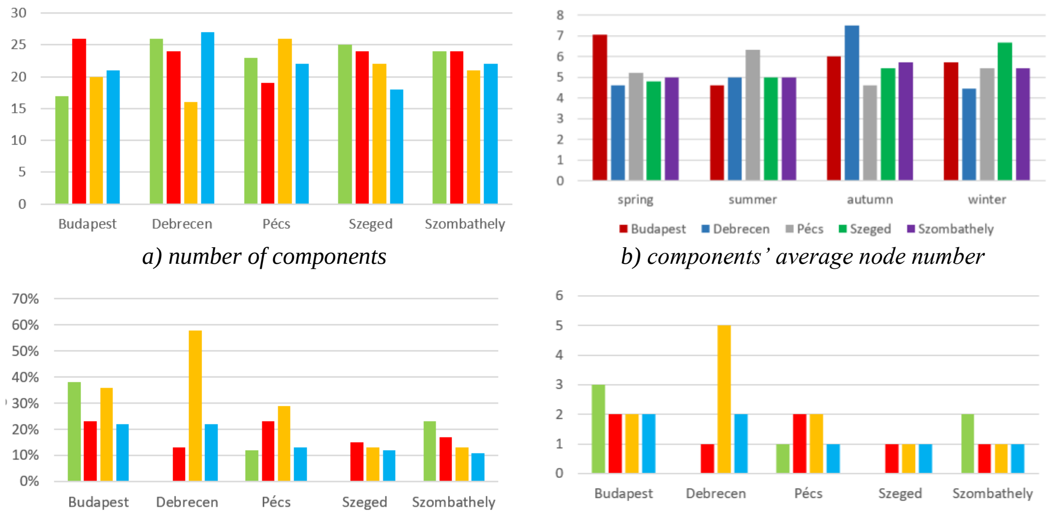


Fig. 5. – Network components based on highest similarity index
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

Analyzing the resulted network, we can observe that the number of components is relatively stable at different municipalities for different seasons, this is particularly true for Szombathely, while more pronounced differences can be observed for Budapest and Debrecen (**Fig. 6.a**). It is also noticeable that, with the exception of Szombathely, the other municipalities seem to be most coherent in different seasons, Budapest in spring, Debrecen in autumn, Pécs in summer and Szeged in winter. Similarly, the highest fragmentation pattern for different seasons appears at different locations: summer at Budapest, winter at Debrecen, autumn at Pécs and spring at Szeged. The above findings are confirmed by the average number of the components (**Fig. 6.b**).

The number of large components and the ratio of the number of nodes in each of them divided by the total number of nodes shows a similar pattern (**Fig. 6.c,d**). It's important, however, that in the case of Debrecen and Szeged, no large components could be identified for spring season, which means that this season shows the highest degree of heterogeneity, with the fewest number of years with similar weather patterns. Although, with varying degree, autumn shows the greatest cohesion for Budapest, Debrecen and Pécs, as significant number nodes are held by the large components of this season.



a) number of components *b) components' average node number*

c) percentage of nodes in large components *d) number of large components*

Fig. 6. – Evaluation graphs on number of network components and number of nodes
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

The visual interpretation of the third data model containing only the edges necessary for all shortest paths is impossible due to the complexity of the network (**Fig. 7**). Its analysis can only be based on network indicators. In this case a separate network model was create for every season and every municipality. Four common characteristics of networks were used in data evaluation (Newman, 2010): the diameter of the network, the weighted diameter of the network, the density of network connections, and the total cost of the network divided by the number of connections.

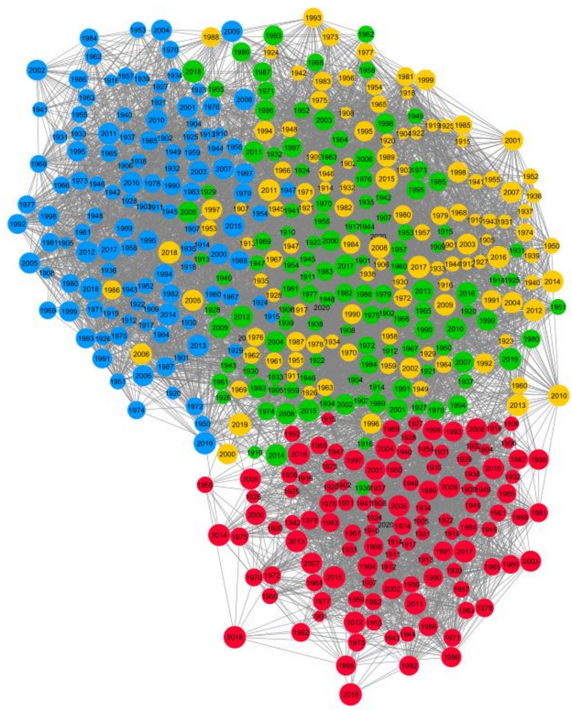


Fig. 7. –Result obtained using third network data model – case of Szeged

The diameter of a network is the longest shortest path that connects node pairs. The smaller the diameter value, the node pairs are closer to each other, which means that the global rate of change is smaller. If we look at the weighted diameter, which shows the highest weight of the shortest paths, we can observe that in some cases these values roughly follow the variation in network diameter (**Fig. 8**). For Budapest and Szeged this is the case for all seasons, but in case of Pécs, for example, although the diameter of the summer network model is the smallest, the weighted diameter value is the largest. Similarly, for Pécs, the network diameter of the autumn season is the largest, yet the weighted diameter is the smallest. If the two indicators do not vary in the same direction means that there are routes with intermediate nodes, resulting shorter lengths than those links which connects directly the two nodes. We can conclude that the two nodes are distant in terms of similarity. This is the case for the summer at Pécs and partly Debrecen.

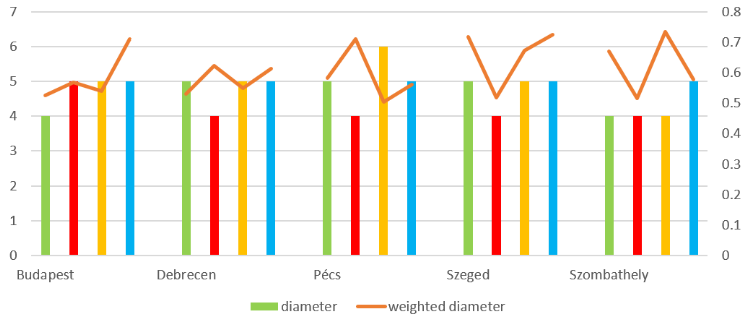


Fig. 8 – Diameter and weighted diameter of networks based on third network model
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

For the same diameter values, if the weighted diameter is higher, this indicates a higher divergence, indicating greater seasonal variability. Thus, for example, in case of summer for Debrecen, Pécs, Szeged and Szombathely (having identical diameter), the highest variability has Debrecen. The diameter of winter is the same for all the municipalities, but the weighted average shows the strongest variation for Budapest and Szeged. Autumn and spring also have highest variation for Szeged than in case of other municipalities.

The connection density of the network model refers to how many of all possible connections was kept in order to have all shortest paths in the structure. The lower this value is, the fewer 'alternative' routes are needed between two nodes, more frequently they can be connected at the lowest cost using direct connections, and hence the lower the variability is within a given season. From **Fig. 9** which plots these densities, we can see that for all five municipalities, winter is the least variable and summer the most variable, even though there is no significant difference between the density indices for spring, summer and autumn at Pécs, Szeged and Szombathely.

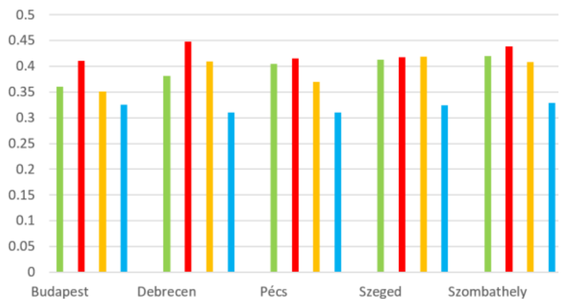


Fig. 9 – Connection density values
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

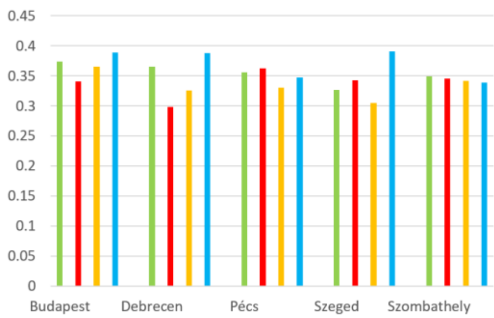


Fig. 10 – Seasons cohesion based on average cost distance (higher value for lower cohesion)
(colors meaning – winter:blue; spring: green; summer: red; autumn: yellow)

The last indicator of the third network model shows the total cost of the network divided by the number of connections. Since the network is constructed based on the differences between the seasons, a higher average value per connection indicates a higher average difference. This shows (**Fig. 10**) that the difference between years is largest for winter, with the lowest average cohesion at Budapest, Debrecen and Szeged. However, the values do not differ that much, that individual comparison can be useful, and the values were included in three cohesion categories. In low cohesion category belongs the spring, autumn and winter in case at Budapest, the spring and winter at Debrecen and the winter at Szeged. The medium category includes the summer at Budapest, all seasons at Pécs and Szombathely, and the autumn at Debrecen. The summer at Debrecen and the autumn at Szeged have high cohesion, they being the most stable seasons.

4. CONCLUSIONS

The aim of this research was to investigate how different network data models based on weather parameters can be used to model and assess climate changing. Five municipalities from Hungary were selected as test locations. Network modelling is a novel approach to climate studies with no specific methodological literature. This research was an attempt to explore the possibility to link two fields: networks and climate. Through the research several network data models were envisioned, but three of them were actually used. One of the important conclusions of this study is that different data models provide different opportunities for data mining. This opens the possibility of further researches to develop other useful network data models. Different data models, as they use different kind of structure and building rules, can highlight different aspects of the changing climate. Admitting this, it can be seen that the individual results do not always converge in the same direction with the same intensity, but the totality them can still emphasize different changes. Since climate change itself is multifaceted, manifesting itself in different ways in different locations, different data models and their indicators can capture different elements of the change.

One of the results of my research was actually the construction of the models, for which I created several modules and functions developed in R Cran. Finding programming methods that could do this in a time-efficient way was very important in the case of large amounts of data processing.

Interpreting a significant number of results for a single municipality or for a single season is a complex task. Two different examples of it are highlighted for Budapest and Debrecen.

Winter at Budapest is changing in many ways, supported by all three models:

- form the first network model, it can be seen that 78% incorporation into other seasons, with the last 30 years mostly incorporation into spring
- form the second network model, tis observable that it has many components, showing considerable variation, with the proportion of large components being the lowest compared to other seasons at that location
- from the third network model the changing characteristic of the winter is shown by the highest weighted diameter and it has the lowest cohesiveness

Autumn at Debrecen is a stable season, based on the following considerations:

- according to the first network model, the autumn compact does not appear in the branches of other seasons
- the second network model shows a high number of large components but a low number of components compared to other seasons, both indicating a compact situation. This is also confirmed by the average number of component nodes, which is the highest for all municipalities
- the third network model shows a low weighted diameter of the network, with a cohesion value that is the second strongest of the four seasons

The performed research confirms that network data models can play a role in climate change modelling and that they have a place in this field, not necessarily stand-alone, but as a valuable complementary method. It is a novel data mining approach, in which the structural characteristics of the network data models are used in analysis, the weather data being used just in the models building phase.

ACKNOWLEDGEMENTS

The presented research was supported by the DOMUS scholarship program of the Hungarian Academy of Sciences.

REFERENCES

Arquilla and D. Ronfeldt. *Networks and Netwars: The Future of Terror, Crime, and Militancy*. RAND: Santa Monica, CA, 2001.

Borgs, C., Chayes, J., Daskalikis, C., Roch, S. (2007), First to market is not everything: an analysis of preferential attachment with fitness, *STOC'07 Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, 135-144, DOI: 10.1145/1250790.1250812

Bulkeley, H. (2012). *Cities and Climate Change*. Routledge. DOI: 10.4324/9780203077207

Csardi G, Nepusz T (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems, 1695. https://igraph.org.

Emmert-Streib F., Tripathi S., Yli-Harja O., Dehmer M. (2018), Understanding the World Economy in Terms of Networks: A Survey of Data-Based Network Science Approaches on Economic Networks, *Frontiers in Applied Mathematics and Statistics*, 4, DOI: 10.3389/fams.2018.00037

Han. J., Kamber, M., Pei, J. (2012), *Data Mining: Concepts and Techniques*, Elsevier, DOI: 10.1016/C2009-0-61819-5

Hunt, A., Watkiss, P. Climate change impacts and adaptation in cities: a review of the literature. *Climatic Change* 104, 13–49 (2011). DOI: 10.1007/s10584-010-9975-6

Ihaka, R., Gentlman, R. (1996), R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*. 5 (3): 299–314. DOI:10.2307/1390807.

Hopkins, L. (2007), Network Pharmacology. *Nature Biotechnology*, 25: 1110-1111,

Li, C., Maini, P.K. (2005), An evolving network model with community structure, *Journal of Physics: a mathematical and general*, 38 (45), 9741-9749 DOI: 10.1088/0305-4470/38/45/002

Light, S., Kraulis, P., Elofsson, A. (2005), Preferential attachment in the evolution of metabolic networks, *BMS Genomics*, 6,159, DOI: 10.1186/1471-2164-6-159;

Magyari-Sáska Zs., Dombay S (2020), Seasons' shifts in some depressions of the Eastern Carpathians, based on daily temperature analysis, *Air and Water Components of the Environment*, Conference proceedings, 213-222

Magyari-Sáska, Zs. (2019), Road Network Based Community Detection. Case Study for an Eastern Region of Austro-Hungarian Monarchy, *Geographia Technica*, 14(1), 82-91, DOI: 10.21163/GT_2019.141.06

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*. 37 (1): 17-23. DOI:10.2307/2332142

Newman, M. (2010), *Networks: an Introduction*, Oxford University Press

Probáld F. (2014), The urban climate of Budapest: past, present and future, *Hungarian Geographical Bulletin* 63 (1) (2014) 69–79. DOI: 10.15201/hungeobull.63.1.6

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504.

Stone, B. (2012) The City and the Coming Climate. *Climate Change in the Cities we Live*. Cambridge, Cambridge Univ. Press, 198 p.

Tan PN, Steinbach M, Kumar V (2005). *Introduction to Data Mining*. ISBN 0-321-32136-7

Thompson D.J., (2009), Shifts in season, *Nature* 457, p 391-392