

PREFERENTIAL ATTACHMENT IN MODELING CLIMATE CHANGES. TEST LOCATION: MIERCUREA-CIUC, ROMANIA

Zsolt MAGYARI-SÁSKA¹, Ștefan DOMBAY¹

DOI: 10.24193/AWC2021_15

ABSTRACT. Climate change is a fact that we face year after year. Although is a common syntagma its manifestation is different for various region of the planet producing not just global, but local anomalies and changes. In order to track these changes, we propose a network model with preferential attachment, vertices representing successive time periods. The test location for our research was Miercurea Ciuc, one of the coldest locations of Romania. We have developed a similarity index including different meteorological parameters such as air temperature, ground temperature, precipitation amount, snow depth and sunshine hours. Using this similarity index for preferential attachment and considering the appearance order of nodes representing periods on time scale we have created a network model which shows the similarities between these periods as they appear in time. Clustering the obtained graph model, we could observe that the created network model at monthly scale clearly shows up some of experienced characteristic at the study location.

Keywords: network model, similarity, eccentricity, Romania, R script

1. INTRODUCTION

Although climate change is evident and does not need to be proved repeatedly, its effect may vary by location. These local manifestations should be analyzed in accordance with the local characteristic (Melo et. al., 2013; Meresa et. al, 2017). These changes may be noticed both at the level of extreme phenomena, extreme values appearance or frequency (Magyari-Sáska and Dombay, 2017) but also at the level of mean values in some cases resulting in season shifts (Magyari-Sáska and Dombay, 2020; Thompson, 2009).

The basins in the Eastern Carpathians always had a specific climate with severe winters and moderate summers (Micu et. al, 2014; Petres et. al, 2017). The perception of unchanging climate could not resist for a long time nor in these isolated regions. In some cases, the change's effects were noticeable sooner by ordinary people, than the detection of changes in climate variables.

This research tries to investigate a possibility to highlight and track climate change using a network model which eventually can be used as complementary analysis beside the existing statistical investigations.

¹ Babeș-Bolyai University, Faculty of Geography, Gheorgheni Extension, 535500 Gheorgheni, Grădina Csiky, zsolt.magyari@ubbcluj.ro , stefan.dombay@ubbcluj.ro

The main idea was to adopt the preferential attachment for more or less aggregated time moments such as days or months, these representing the vertices of the network. Preferential attachment has been used at a large scale in various domains (Li and Maini, 2005; Light et al., 2005; Borgs et. al, 2007), it has not however been utilized in climate applications. It reflects the fact that a newly appearing vertex in a network doesn't connect arbitrarily to the existing ones, rather there is a preference for some of them (Barabási and Albert, 1999). In our case this preference was ruled by the climatic similarity between the two vertices.

Our assumption was that for a quasi-stationary climate in the obtained network the common temporal periods (seasons, months) should be clearly visible with less mixtures between them and there should not be long tails inside them. To investigate these characteristics, we used specialized indicator used in network analysis and graph theory.

What we knew from local experiences was that winter has changed a lot having higher temperatures and much less snow and also summer has changed having less compact warm period. We wanted to see if these changes are observable in network representation, if it's about a shift or mixture between months or it's a dynamic evolution of climate.

2. DATA AND METHODS

The data we used were collected from the land based meteorological station at Miercurea-Ciuc, including the following parameters at daily resolution: mean air temperature, mean ground temperature, minimum and maximum air temperatures, relative air humidity, snow depth, sunshine hours, precipitation amount, wind direction and average wind speed. Data covers the 1999-2019 years inclusively.

As mentioned before for analysis a network model was adopted in which the vertices were data values represented on a monthly scale while the appearance of connections was based on preferential attachment.

To achieve the preferential attachment, we used two simple rules:

- the appearance of a new vertex in the network is based on its show up in time
- a new vertex, representing a month in a year, will be attached to that existing vertex in network to which is closer considering their aggregate meteorological parameters. Depending on our decision it was possible to have multiple connections for a vertex based on a threshold value.

Due to scale differences at different parameters at the beginning we normalized them using equation 1. Later we used these values to calculate the closeness grade between vertices.

$$n_i = \frac{min-v_i}{max-min} \quad [1]$$

where,

n_i – the normalized value
 v_i – the original value

min, max – the minimum and the maximum values respectively for the whole data set

To prepare the network model described previously a function was developed in R (Fig. 1), which has scripting and data analyzing capabilities, becoming a powerful tool in various statistical and data handling applications even from its beginning (Ihaka and Getlman, 1996). The script takes the data source, defined as a CSV file and outputs the nodes and edges of the built network in such format that it could be imported in major network analyzing software. The script also has options to define the starting and ending year, a filter system to select and weight which meteorological parameters should be used to calculate the closeness value and also a switch to determine whether a new vertex can be connected to one (the closest one) or to multiple existing vertices based on the threshold value. Applying a filter on the used meteorological parameters allows to choose their different combinations and to study the related influences on the resulted network architecture.

```

graph <- function(name,y1,y2,w,type=0)
{
  d = read.csv(paste(name,"csv",sep="."))
  rlink = data.frame(
    Source = c(0),
    Target = c(0),
    diff = c(0),
    timecost = c(0))
  rnode = data.frame(
    ID = c(0),
    year = c(0),
    month = c(0))
  l = length(d[,1])
  nm=normalize(d,w)
  lb=min(which(d$year==y1))
  hb=max(which(d$year==y2))
  for (i in (lb+1):hb) {
    rnode = rbind(rnode, c(i, d$year[i], d$month[i]))
    p=mindist(i,lb,nm,0.001)
    t2 = as.Date(paste(d$year[i], "-", d$month[i], "-", d$day[i],
sep = ""))
    lp = length(p)
    if (type==1) lp=2
    for (j in 2:lp) {
      t1 = as.Date(paste(d$year[p[j]], "-", d$month[p[j]], "-",
d$day[p[j]], sep = ""))
      rlink = rbind(rlink, c(i, p[j], p[1], t2 -
t1))
    } }
  rnode = rnode[-1, ]
  rlink = rlink[-1, ]
  write.csv(rnode, "nodes.csv", row.names = FALSE)
  write.csv(rlink, "edges.csv", row.names = FALSE)
}

```

Fig. 1. The main R script to create the network model

The resulted file memorizing the vertices holds the year and month of every element, while the file with the connections contains beside the source and target

vertices the normalized difference between them and also the time shift between those vertices expressed in days.

Gephi software was used to evaluate the resulted networks. It offers not just a visualization for complex networks but some basic and also advanced network analysis. From the offered options we needed the vertices degree, diameter and also eccentricity as basic indicators in network analysis (Harary, 1969; Newman, 2010).

Degree of a vertex is defined as the number of edges relate to it, which in case of directed networks may be divided in in-degree, counting the incoming connection to that vertex and the out-degree representing to outgoing connections from a vertex (Fig. 2). Our network is surely a directed one as the linkage between vertices can happen only in one direction. A newer vertex is linked to an existing one and opposite direction linking has no meaning in the network construction phase.

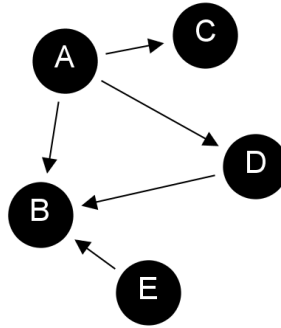


Fig. 2. In-degree and out-degree of a vertex
(A with the highest out-degree; B with the highest in-degree)

In general, assuming preferential attachment the degree of a vertex marks how important, attractive may the vertex be in the network. The case of a higher in-degree show a more characteristic day or month and help to evaluate the stationarity or dynamism (evolution) of the climate changing process.

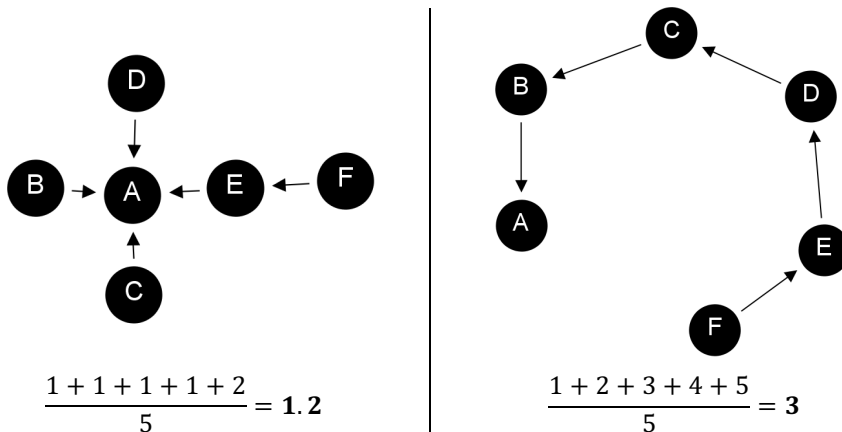


Fig. 3. High and low mean eccentricity against vertex A

The diameter of a network is by hdefined as the length of the longest path between any vertex pair. In our case, considering direction between connected vertices, as longest a path is, as probable to face a dynamically evolving process rather than a stationary one, having a higher elongation in the network

In our case the diameter of a network is practically the highest eccentricity value. Eccentricity is defined for every vertex and represents the longest path from that vertex to any other vertex in the network. Mean eccentricity for a network may be a better characteristic for elongation than diameter (Fig. 3).

Considering our network architecture this eccentricity value leads to the first vertex of the network. Surely this value cannot be a good characteristic for a vertex because the path may pass through a huge number of vertices, and as much as we depart from the original vertex as much it becomes more irrelevant. To counteract this problem, we started to use subnetworks.

A subnetwork or a subgraph as defined is a subset of vertices from the original network keeping all connections between the remaining vertices. The eccentricity of every vertex was considered only to the point of origin or other selected vertex of the subnetwork (Fig. 4). In cases when not the point of origin is considered the direction of edges are disregarded. In cases when a subnetwork defines a well delimited category this kind of limited eccentricity may be useful.

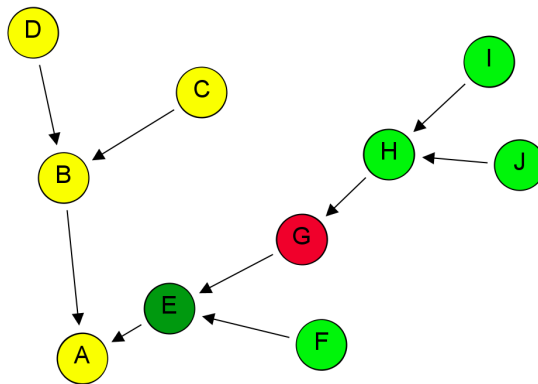


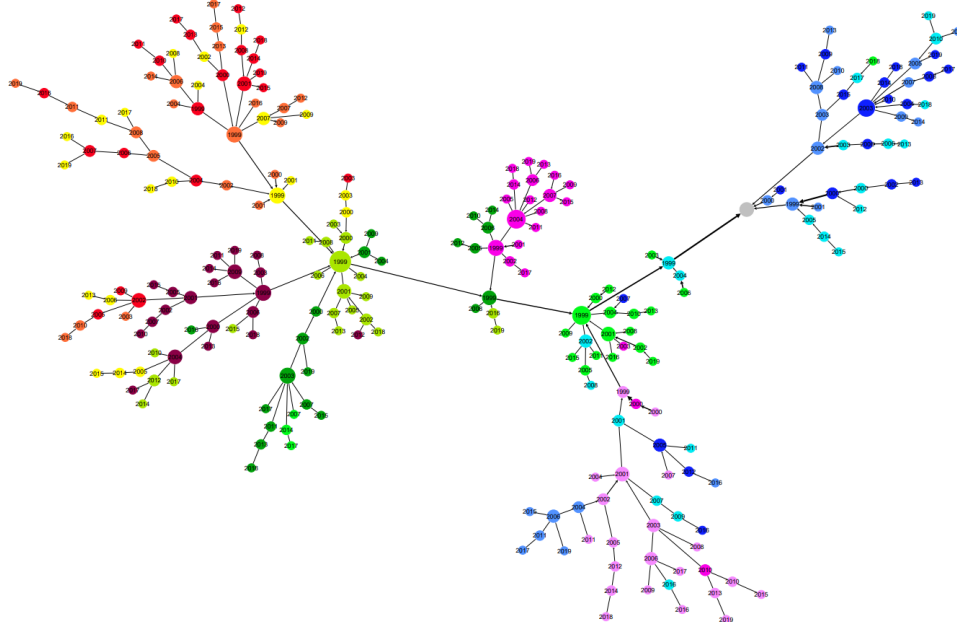
Fig. 4. Subnetwork with point of origin in E; limited eccentricity of 1.6 against vertex G

3. RESULTS AND DISCUSSION

Initially we wanted to construct and evaluate the network at different time scales. Starting with daily resolution the resulted an extremely complex network with over 4000 nodes and connections. That network could not be visually inspected and the obtained data does not reveal any specific elements and no patterns could be identified.

Considering that data aggregation may help we calculated monthly scale averaged value for every meteorological parameter, resulting in visually operable network with just 132 nodes. After setting a coloration based on month and vertices size became proportional with their in-degree we obtained what is shown in figure 5.

At first sight it has some kind of homogeneity, a pattern can be identified based on colors, but for systematical analysis it we have to consider by parts. The “bone” of the network is distinguishable as the first months from 1999 logically have to connect to the previous month. But after that the structure of the network starts to became more different which made necessary an analysis by parts based on a common investigation method.



**Fig. 5. The resulted network model at monthly scale
(same color identifies same months from different years)**

As a color (month) based grouping was observable we tried to characterize each month category (values for different years but the same months) separate. To do this we had to develop a method to determine where a subnetwork corresponding to a given month begins, where is its point of origin.

We applied the following idea: for every month category we selected the most relevant vertex, that with the highest in-degree, considering that it characterizes the most that month. To do not omit vertices that appears before the relevant vertex we have followed the connections starting from the relevant vertex till we have met a vertex representing a different month type. The last vertex of this path was considered the origin point for a month type.

For example, the most eloquent situation can be found for October (Fig. 6). The characteristic vertex is the one from 2004, having an in-degree of 7, meaning that from the following period 7 years had a most similar month with it. The last vertex showing October in the path towards the first vertex of the network is 1999, which became the point of origin for October months. To characterize the resulted subnetwork, we have the maximum and the mean eccentricity of it, but as there are vertices indicating other types of month in the subgraph, we introduced two

concepts: completeness and homogeneity of the subnetwork. Completeness is percentual expression of the number of months of a given type against the total number of months with the same type. Homogeneity is also a percentual expression presenting the number of months of a given type against to the total number of months inside the selected subnetwork.

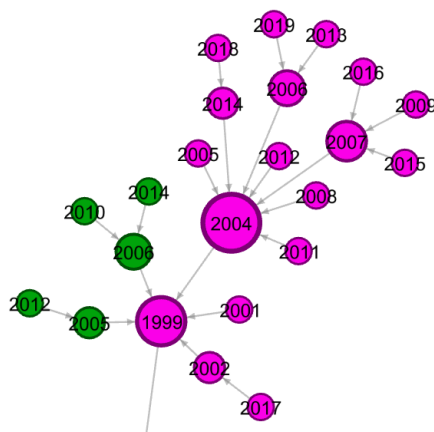


Fig. 6. Subnetwork for month of October

For October months, the completeness is 86%, the homogeneity is 78%, the maximum eccentricity is 3, and the average eccentricity is 2.05. This last value indicates that the similarity of a month of October to point of origin of the subnetwork (1999) is at 2 steps away. If we consider the reference year the most characteristic node of the subgraph (in this case the year 2004), the maximum eccentricity remains 3, but the average decreases to 1.57. The average eccentricity calculated for the central vertex will be call specific average eccentricity. Based on this latter value we can say that the months of October were very similar and there is no trend of elongation in this subgraph. To quantify the elongation, the specific average eccentricity can be reported to the number of nodes which in this case is 0.04.

Trying to apply the described method we observed some strange situations. While for the majority of months there was no problem identifying a central vertex, for two of them – August and December – there were two central nodes (with the highest in-degree value) in two different places in the network. In fact, these months were spread over the network doesn't present a coagulated part, they were dissolved. Observing this we slightly refined the calculation methods for every subcomponent and in those cases in which a dissolved component was present the vertices of them were excluded from the calculations for the containing subnetwork.

After identifying the subnetworks appeared also a visual representation of the months (Fig. 7), which reveals another anomaly. February is completely dissolved in the network, there is no central vertex for it, we could not grab to say: from this point begins at least a part for months of February.

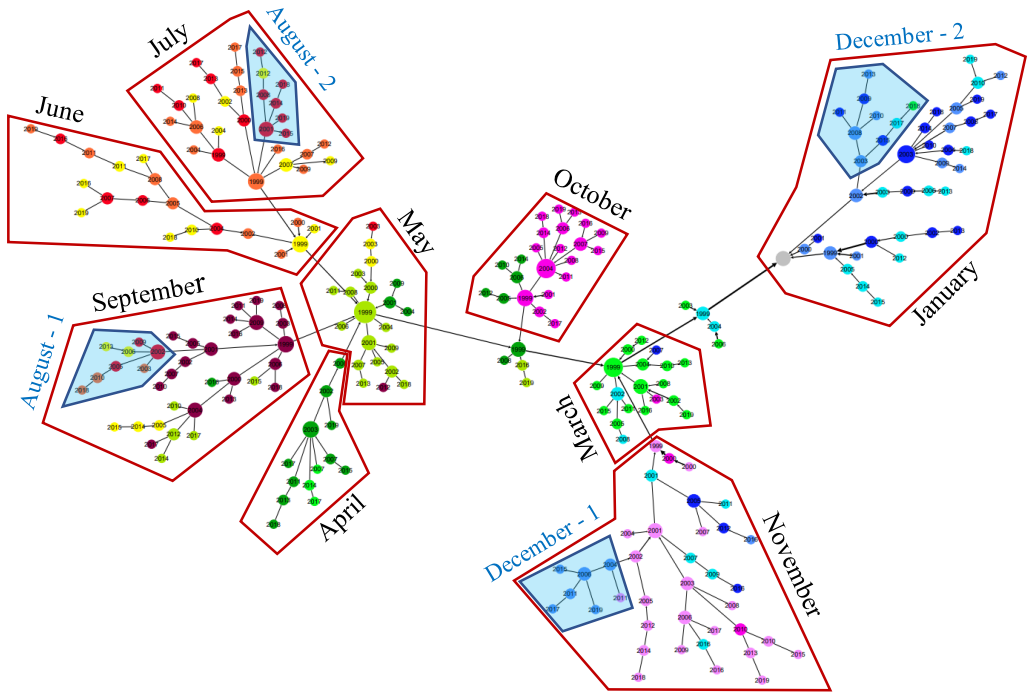


Fig. 7. Identified subnetworks with their corresponding months based on central node

Continuing the analysis for all identifiable subnetworks we made the calculation separately for the two part in case of August and December and we omitted February. The results are summarized in table 1. As well from the interpretation of values August and December may be excluded as the already represent splitted elements, doesn't form a compact region to characterize the evolution of one month.

Table 1. Subnetwork characteristics for different months as shown in figure 7

Month	Central vertex	Homogeneity	Completeness	Maximum eccentricity	Specific eccentricity	Specific elongation	Total elongation
January	2003	41%	67%	7	3.16	0.09	0.21
March	1999	79%	71%	3	1.73	0.09	0.16
April	2003	77%	48%	3	1.50	0.12	0.23
May	1999	67%	67%	3	1.64	0.08	0.14
June	1999	43%	38%	6	3.75	0.20	0.32
July	1999	50%	52%	4	2.18	0.10	0.18
August 1	2002	38%	14%	1	1.00	0.13	0.13
August 2	2001	88%	33%	3	1.28	0.16	0.38
September	1999	53%	95%	4	1.95	0.05	0.11
October	2004	78%	85%	3	1.05	0.05	0.13
November	2001	95%	95%	5	2.70	0.09	0.16
December 1	2006	86%	29%	2	1.00	0.14	0.29
December 2	2008	44%	19%	2	1.00	0.11	0.22

Each value from the table can interpreted in terms of climate change. Months where the homogeneity value is a low score may not be considered as a clearly defined time period with specific characteristics that doesn't alter with other months. This is the case of January, September and especially June. Whether this lack of

homogeneity is due to “contamination” from other months or to “migration” of months to other locations in the network, it may be characterized by the completeness value. Considering both mentioned values March, May, October and November may be considered stable months, while April not, even if it has a high homogeneity value its completeness is low, as important number of vertices (months) are not present in the subnetwork.

Another important element is the dynamism inside the subnetworks which may be characterized by the elongation, which is a value calculated from eccentricity. Specific elongation takes into account the specific eccentricity which is based on the central vertex of the subnetwork, while total elongation is based on maximum eccentricity and therefor consider the point of origin of the subnetwork. Based on this latter March and November is considered a more changing month than May or October.

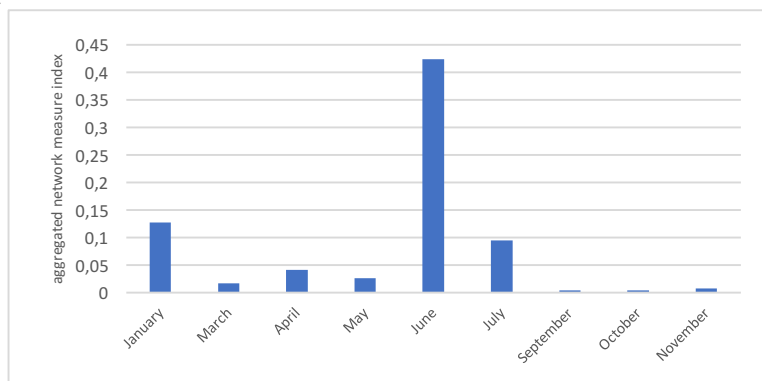


Fig. 8. Aggregated network measures to characterize changes of months (higher values – higher change)

In order to facilitate the interpretation, we have tried to summarize these values and to have a comprehensive characterization of what happens in every month type, we aggregated the homogeneity, completeness, specific eccentricity and maximum elongation values (Fig. 8). Lower values represent higher stability in climate for that month, while higher values a more dynamic evolution as we described previously.

4. CONCLUSIONS

Our study aimed to offer a new identification method of climate change using a network model built on preferential attachment. Through the described and applied methodology on a monthly scale, the temporal evolution and characteristics of climate can be clearly identified. Using and combining the proper network indicators both the anomalies (dissolved months, emphasized changing in month characteristics) and stability (months with slow changes) can be identified.

During the research we not just used some existing network measures but also had to define others in order to can quantify the visually observed results. The development and existence of these kind of indicators are important because in complex networks visual identification cannot be made.

At the moment we had a limited number of data available for us and we tried to integrate all of them. Using a larger dataset and working with specific and calculated values as inputs the proposed model could and should be evaluated in climate change impact for different sectors.

For our study site we identified stable spring and autumn, beside rapidly changing summer and a very heterogenous winter for which December and February are not outlined at all. These results are confirmed by those who lives in this region. Using all available meteorological parameters, the months of spring and autumn are well separated in the network, a situation which doesn't happen if we consider only temperatures. This research till this moment used all available meteorological parameters but surely it worth to be considered separately or grouped which may reveal other aspects of the changing climate. However, we consider the presented method a new approach that can contribute in studying the changing climate both globally and locally.

REFERENCES

1. Barabási, A.-L., Albert, R. (1999), Emergence of scaling in random networks. *Science*. 286 (5439): 509–512. doi:10.1126/science.286.5439.509
2. Borgs, C., Chayes, J., Daskalakis, 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: <https://doi.org/10.1145/1250790.1250812>
3. Harary, F. (1969), *Graph Theory*, Addison-Wesley, 1969
4. Ihaka, R., Gentleman, R. (1996), R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics. American Statistical Association*. 5 (3): 299–314. doi:10.2307/1390807.
5. 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: <https://doi.org/10.1088/0305-4470/38/45/002>.
6. Light, S., Kraulis, P., Elofsson, A. (2005), Preferential attachment in the evolution of metabolic networks, *BMS Genomics*, 6,159,doi:<https://doi.org/10.1186/1471-2164-6-159>
7. 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
8. Magyari-Sáska Zs., Dombay S (2017), Changes in Europe's Temperature Extreme Values., *Air and Water Components of the Environment, Conference proceedings*, 77-84
9. Meresa, H.K., Romanowicz, R.J. & Napiorkowski, Understanding changes and trends in projected hydroclimatic indices in selected Norwegian and Polish catchments, J.J. *Acta Geophys.* (2017) 65: 829. <https://doi.org/10.1007/s11600-017-0062-5>
10. Micu D.M., Dumitrescu A., Cheval S., Birsan M-V, (2014), *Climate of the Romanian Carpathians: Variability and Trends*, Springer
11. Newman, M. (2010), *Networks: an Introduction*, Oxford University Press
12. Petres, S.; Korodi, A.; Keresztes, R.; Szep, R. (2017), Tendencies and Particularities in Thermic Inversion Episodes in the Ciuc Basin - Eastern Carpathians, Romania, *Conference Proceedings, 17th International Multidisciplinary Scientific GeoConference SGEM 2017*, 17(14), 445-452
13. Thompson D.J., 2009, Shifts in season, *Nature* 457, p 391-392