UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Utilizing Self-organizing maps to determine animal dominance

Lucas Esperancini Moreira e Moreira

Monograph - MBA in Artificial Intelligence and Big Data

ICMC USP
SÃO CARLOS

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________________

Lucas Esperancini Moreira e Moreira

Utilizing Self-organizing maps to determine animal dominance

Monograph presented to the Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, as part of the requirements for obtaining the title of Specialist in Artificial Intelligence and Big Data.

Concentration area: Artificial Intelligence and Big Data

Advisor: Prof. Dr. Ricardo Cerri

Versão original

São Carlos

2023

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m

Esperancini, Lucas M M
Utilizing Self-organizing maps to determine animal dominance / Lucas Esperancini Moreira e Moreira ; orientador Ricardo Cerri. – São Carlos, 2023.
49 p. : il. (algumas color.) ; 30 cm.

Monograph (MBA in Artificial Intelligence and Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2023.

1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Dissertação. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. Cerri, Ricardo orient. II. Título.

Lucas Esperancini Moreira e Moreira

Utilizing Self-organizing maps to determine animal dominance

Monografia apresentada ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial e Big Data

Advisor: Prof. Dr. Ricardo Cerri

Original version

São Carlos

2023

Folha de aprovação em conformidade
com o padrão definido
pela Unidade.
No presente modelo consta como
folhadeaprovacao.pdf

"I'm old enough to know you can't close your mind to new ideas. You have to test out every possibility if you want something new."

John B. Goodenough

ABSTRACT

Esperancini, Lucas M M **Utilizing Self-organizing maps to determine animal dominance**. 2023. 49p. Monograph (MBA in Artificial Intelligence and Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

The animal dominance in a group can interfere on how much access each individual has to a limited resource, for instance, food or water. Therefore, a possible technique to increase the weight gain in a social group, for example in a livestock farm, would be to determine the social relation amongst these individuals. This paper proposes a method to help determine the social dominance of animals using geo-spatial data and self organizing maps (SOM). Self organizing map is an unsupervised machine learning technique that aims to identify similar data, in this case, animals that are more prone to be dominant or submissive. A variant of this method was used on a data set of 36 white-bearded wildebeest from a study in Kenya. These animals were attached to a GPS collar that was programmed to collect 16 points during the day during three year. The data was separated monthly in order to apply the SOM. In general, the results of the proposed method managed to separate the animals in different hierarchies in some months. However, with more data from the animals and with better understanding from the enviroment that the animals are in could provide more promising results.

Keywords: Dominance. SOM. Hierarchy. Self-organizing map. Cattle.

RESUMO

Esperancini, Lucas M M **Utilizing Self-organizing maps to determine animal dominance**. 2023. 49p. Monografia (MBA em Inteligência Artificial e Big Data) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2023.

A dominância animal em um grupo pode interferir no acesso que cada indivíduo tem a um recurso limitado, por exemplo, comida ou água. Portanto, uma possível técnica para aumentar o ganho de peso em um grupo social, por exemplo na pecuária, seria determinar a relação social entre esses indivíduos. Este trabalho propõe um método para ajudar a determinar a dominância social dos animais usando dados geoespaciais e *Self-organizing maps* (SOM). O SOM é uma técnica de aprendizado de máquina não supervisionado que visa identificar dados semelhantes, neste caso, animais mais propensos a serem dominantes ou submissos. A variante deste método foi usada em um conjunto de dados de 36 gnus de barba branca de um estudo no Quênia. Esses animais foram presos a uma coleira GPS programada para coletar 16 pontos durante o dia durante três anos. Os dados foram separados mensalmente para aplicar o SOM. Em geral, os resultados do método proposto conseguiram separar os animais em diferentes hierarquias em alguns meses. No entanto, com mais dados do animais e com uma melhor compreensão do ambiente em que os animais estão poderiam fornecer resultados mais promissores.

Palavras-chave: Dominância. SOM. Self-organizing map. Gado.

LIST OF FIGURES

LIST OF TABLES

LIST OF FRAMES

LIST OF ABBREVIATIONS AND ACRONYMS

SOM	Self-organizing map

CONTENTS

1 INTRODUCTION

The animal dominance in a group can interfere on how much access each individual has to a limited resource, for instance, food or water. Specially for the beef industry, for this reason, farm managers can use this information to improve animal welfare. Social dominance is recognized as the most important component of social behavior by Syme and Syme (1979) in their review of social structure in farm animals. This dominance can influence how animals are aligned with each other, how close to the center of herd they stay, and how they play with each other. According to Price (2008), these behavior patterns can impact beef production as their access to resources (food, water, shelter) is differentiated by the social hierarchy.

In the review by Hubbard, Foster and Daigle (2021), the authors map many attributes in the literature that were used to pursue the understanding of the social dynamics in a herd like age, sex, body weight and size. The authors also state that identifying indirect measures of social dominance would develop improved management beef cattle group practices. For this reason we propose the usage of GPS data aid this identification.

One possibility to separate animals dominants from submissive with GPS data is to cluster the data using Kohonen's self-organizing maps (or SOM). SOMs are neural networks that forms a nonlinear projection of a high dimensional data into a low-dimensional grid. The low dimensional grid will try to preserve the relations from the input space presenting patterns of the data that are close. In other words, the grid (which is the output) will cluster the animals dominant and submissive in a 2D map.

And so, this paper proposes to use self-organizing maps with geospatial data to aid identify animals more prominent to be dominant and submissive. The method uses GPS data and the distances to resources with the SOM to generate a map that shows the animals more likely to be dominants and submissive.

1.1 Motivation

Farms in the beef industry can use social dominance information to improve animal welfare and production. In order to understand the dominance relation of a herd, it is necessary to collect data and that can very time consuming or expensive.

It was noted that most of the effort to find the hierarchy in a bovine herd was with human observation, either on site or through video feed. Therefore the authors in this research, noted an oportunity to use data collected automatically and without human supervision. In this research, it is proposed to use GPS data of each individual in the herd.

Another choke point found during the evaluation for this research is the difficulty to find labeled data, i.e., a data set that already contains a training set with the animals already recognized as dominants.

In order to surpass this obstacle, this paper proposes to use a specific type of unsupervised neural network. The self-organizing map (SOM) is particularly suited for this application also because it is not necessary to create a number of groups. Instead, the algorithm will plot in a map the degree of dominance of each animal.

Therefore, the goal of this research is to use geospatial data collected from GPS devices that are very accessible and self-organizing maps in order to have a better understanding of the social hierarchy of a herd. The advantage of such methodology would be to use a non-invasive technique with low human labor to interfere with the quality of the data and a technique that does not need labeled data to achieve a response.

1.2 Research Question and Objectives

To understand the social relations of an animal group is a complex task. These relations may impact the access of each individual to a common resource such as food or water. Therefore we propose a methodology to use localization of each individual and their distance to the common resources in order to extract the dominance information. In other words, we propose to cluster all the animals in a herd using geospatial data. Hence, the research question can be stated as:

“Is it possible to determine the social hierarchy of a herd using geospatial data and self-organizing maps?”

Given the research question, the objectives can be summarized as:

- Implement a self-organizing map applied to geospatial data to cluster the animals in groups of dominants and submissive.
- Apply the algorithm to new data of animals to visualize how the relations may occur with different group of animals.
- Evaluate if the algorithm has presented the degrees of dominance of each animal coherently with their moving patterns

2 DEVELOPMENT

This chapter contains all the sections regarding the development of the current study. First, in the Section 2.1 the definitions that are going to be used in this paper are presented. Next, in the Section 2.2, it is discussed the work related to this research.

2.1 Theoretical Framework

2.1.1 Animal dominance

Animal dominance has been studied for many years and it has a broad concept in literature. It has a range of categories such as: privileged role, winner of a given context, priority of access to resources and others (DREWS, 1993). But in general it can be understood as a relation between at least two individuals which one has a superior gain (of food, water, sexual companion) over another individual, the latter being the submissive.

This paper will focus on the definition which the dominant animal has more control over the resources (in this case, food or water) in a common area. In other words, the submissive animal would have lower priority in the action of obtaining the common resource.

The hierarchy discussed here can also be applied for beef cattle. In general, it was studied to find a relation between the social dominance and cattle attributes, productivity, behavior etc. (HUBBARD; FOSTER; DAIGLE, 2021).

2.1.2 Self-organizing maps

Kohonen's map or Self-organizing map (in short, SOM) is a tool to extract information visually, usually from a 2D grid. It was first developed by Teuvo Kohonen in 1982 (KOHONEN, 2013) and since then, it has been used in many applications.

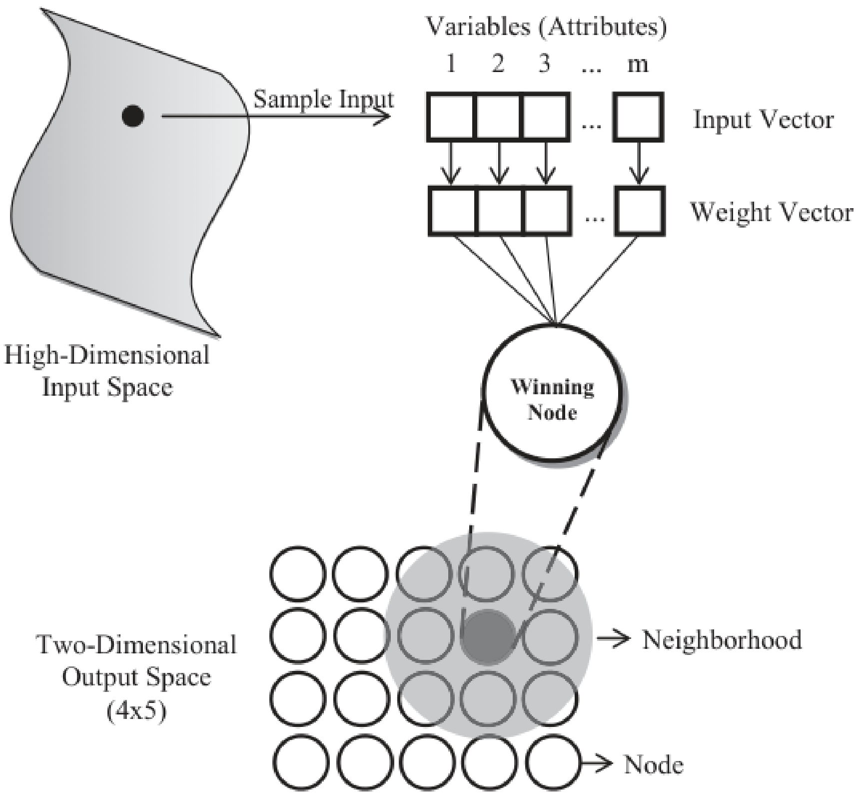
Self-organizing map is an unsupervised learning neural network that reduces the dimension of the data and preserves the relations of the input data into a lower dimensional grid (usually two dimensions). This is particularly useful to have clearer understanding of the data being inputted.

The SOM architecture in its most basic form consists in a neural network of two layers therefore, no hidden layers. First, there is the input layer which is the input data to be described (i.e. all the records) and the output layer, also known as Kohonen's layer. Every node in the output layer is fully connected to every attribute of each record.

The core idea of the algorithm consists of, for every record, calculate the distance with every node's weight and from these distances, it is selected the node from the

Kohonen's layer with the lowest value. This is known as the *best matching unit* (BMU). Then, all nodes within a radius r from the BMU has its weights adjusted to make them more similar to the winning input vector. This process is repeated until a stopping criteria is met. The Figure 1 illustrate one single iteration of the SOM algorithm.

Figure 1 – Representation of the SOM algorithm



Source: Asan and Ercan (2012)

The algorithm can be formally defined as:

- w_{ij}: weight vector associated with the output from row i and column j
- x_k: input k from available data
- d_{kij}: distance from an input k and output node from row i and column j
- r: radius neighborhood
- α: learning rate

For each input record k:

1. Calculate the distance between input k and all nodes on the output layer ($d_{kij} = ||x_k - w_{ik}||$)
2. Encounter the winner node (BMU). The best unit can be found as $w_{\overline{ij}} = min_{i,j \in M}(d_{kij})$
3. Update every node in output layer inside radius r from the winning node $\overline{ij}$ with the update function $w_{ij}(t+1) = w_{ij} + \alpha h(x_k - w_{ij})$.

2.2 State of the art

2.2.1 Animal dominance

According to Drews (1993), social dominance in group animals has many definitions in the literature. Some of them are: privilege role, reproductive status, priority of access and others. This paper, will focus on the priority of access due to its application in beef cattle farms.

One example is from (BANKS *et al.*, 1979), where the authors seek to determine the relation of dominance and access to resources on domestic fowls. In this study the resources were food, water, perch, litter box and nest box. At the end of the experiment it was seen that the social order was maintained through out the experiment and at the food resource was where the most agnostic behavior appeared.

In (BOCCIA; LAUDENSLAGER; REITE, 1988), the experiment was made with macaques of different sex and age. The behavior observations were sparse (twice a week) during the feeding hours. In this study, the ranks did not change along the experiment and animals with high rank reached the food resource first and had access to it for longer periods.

Another example can be seen in (SHINDE; VERMA; SINGH, 2004), where the study was made with goats. In this study the animals in the higher position in the social hierarchy had access to better fodder and they were first to eat in the trough.

A different approach of dominance over food resource can be seen in (CAFAZZO *et al.*, 2010). In this paper, the authors showed that the dominance of dogs during the relations involving food dispute involved only food stealing and withdrawal from the submissive animals.

In general, it could be seen that the study of animal dominance is very diverse. Different methodologies for behavior observations, to determine dominance, food given ad libitum and even number of dominant groups varies. This is showed in Table 1. For a more extensive review, it is recommended (DREWS, 1993) and (HOLEKAMP; STRAUSS, 2016).

However, regarding beef cattle, the study is more narrow but not less important. For example, in (FORIS *et al.*, 2021), the authors use the dominance of animals captured

Table 1 – Summary of studies on animal dominance

Author	Observation period	Session time	Dominant groups	Ad libitum
Banks *et al.* (1979)	App. one month	10 min	2	Yes
Boccia, Laudenslager and Reite (1988)	App. three months	2 hours	2	No
SHINDE, VERMA and SINGH (2004)	ten days	6 hours	2	No
Cafazzo *et al.* (2010)	App. three months	2 hours	2	Yes
Foris *et al.* (2021)	2 years	24 hours	2	Yes
Val-Laillet *et al.* (2008)	two days	24 hours	2	No

Source: Elaborated by the authors

via priority of resource (food and water) to evaluate the usage of a mechanical brush that improves animal welfare.

In (VAL-LAILLET *et al.*, 2008), the authors conclude that high rank cows spent more time at the feeder in comparison to low rank cows. The experiment was made with only cows and the behavior was captured via video cameras for two days.

In general, to determine dominance in cattle, the data was captured visually. This can be seen in the the review by Hubbard, Foster and Daigle (2021). It was found 24 articles that evaluated the social dynamics in herds. In these articles, the social dominance was evaluated via human observation or video capture. In none of these articles covered by the review, it was used GPS data collection.

The only article using GPS data on beef cattle is Weerd *et al.* (2015), however the goal of the study is to identify the action of the animal (foraging, lying, standing or walking). During the research for this paper, no article that used GPS data to determine animal dominance was found.

2.2.2 Self-organizing map

The self-organizing maps or Kohonen's map was first presented by Teuvo Kohonen in 1982 (KOHONEN, 2013). Since then, it has been studied and applied in many fields like: visualization of statistical data and document collections; bio medical applications, for instance in diagnostic methods; data analysis in commerce, industry, macroeconomics, finance (KOHONEN, 1990). However this paper focus only in the articles that have applied SOM with spatial data.

An early application on geo spatial data could be found in (TAKATSUKA, 2001). The authors applied to demographic data (26 attributes) and utilized SOM to analyze the data with the Kohonen's layer being three dimensional. Koua (2003) took a similar approach to use SOM as a framework to extract value from geospatial data. They used the U-matrix distance and also mesh visualization and 3D surface plot for exploratory data analysis and visualization

Bação, Lobo and Painho (2005) proposed to make a variant on the basic SOM. The authors propose to look for best matching unit in two phases. First, it is found the nodes in the Kohonen's layer that best represent the data geographically. Then, the usual approach to find the BMU is done however, only to the nodes found in the first phase.

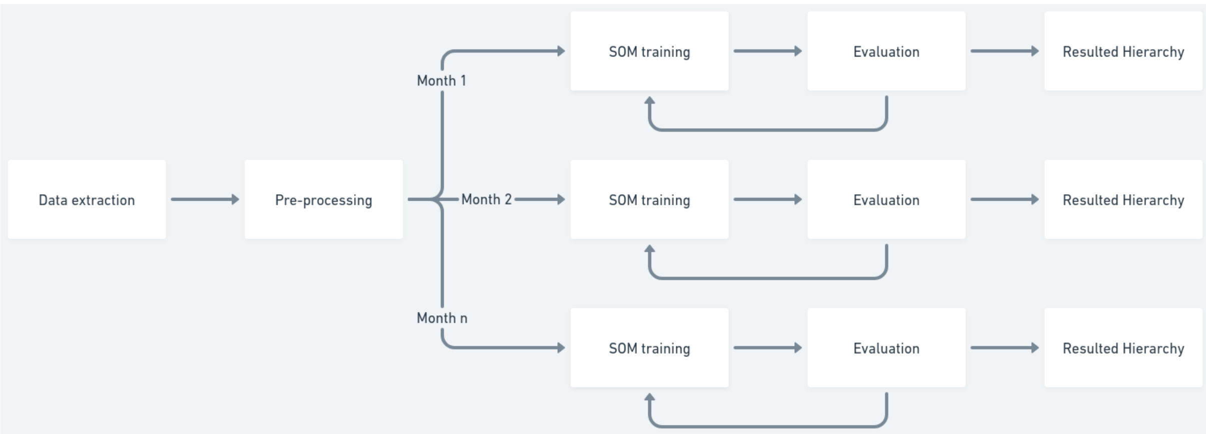
Skupin (2008) presents the usage of SOM with geographic data of moving units through a n-dimensional attribute space. In the chapter, the authors used data mainly from human movement to attempt to determine similar path characteristics. It was applied to determine similar cities across the path taken of each GPS. The second experiment shows how SOM can describe better the social economics characteristic during a destination from point A to point B compared to summarized statistics for example.

2.3 Methodology

This paper proposes the usage SOM to understand the dominance hierarchy in a group of animals. Since SOMs generate relaxed clusters, the hierarchy is not a clear attribute in the animal relationship and during the research, no labeled data was found, SOM was the chosen to be method to attempt to extract this information.

In order to evaluate if SOM was a adequate tool, the methodology proposed is described by Figure 2. This pipeline is described in the following sections.

Figure 2 – Methodology pipeline



Source: from authors

2.3.1 Data extraction

The intention was to use geospatial positioning of the animals, and since it was not in the scope of this study to collect the data described, the approach in this research was to look for open data. The source that most fit the study could be found in https://www.movebank.org/.

This website provides a lot of data related to animal positioning. The data was extracted from the study Stabach *et al.* (2020) and was selected mainly for three reasons:

- In this research, the authors were interested to use data of heavy mammals so this study can be replicated in a reality of cattle in Brazilian farms. The study chosen collected data from wildeebeests.
- The frequency of the data time collection was hourly. This was important in order to obtain more information of the interaction among these animals in an amount of time.
- The last reason being the animals were in relative close position. The animals must be in close position so it is more likely to exist an interaction between them and, therefore, a dominance.

In most studies checked by the authors, the animals were not tracked in proximity or they were not big mammals. Hence, the most adequate data was from Stabach *et al.* (2020).

2.3.2 SOM

The methodology to evaluate the questions proposed, are presented in the next sections. The Figure 3 illustrates what is further described.

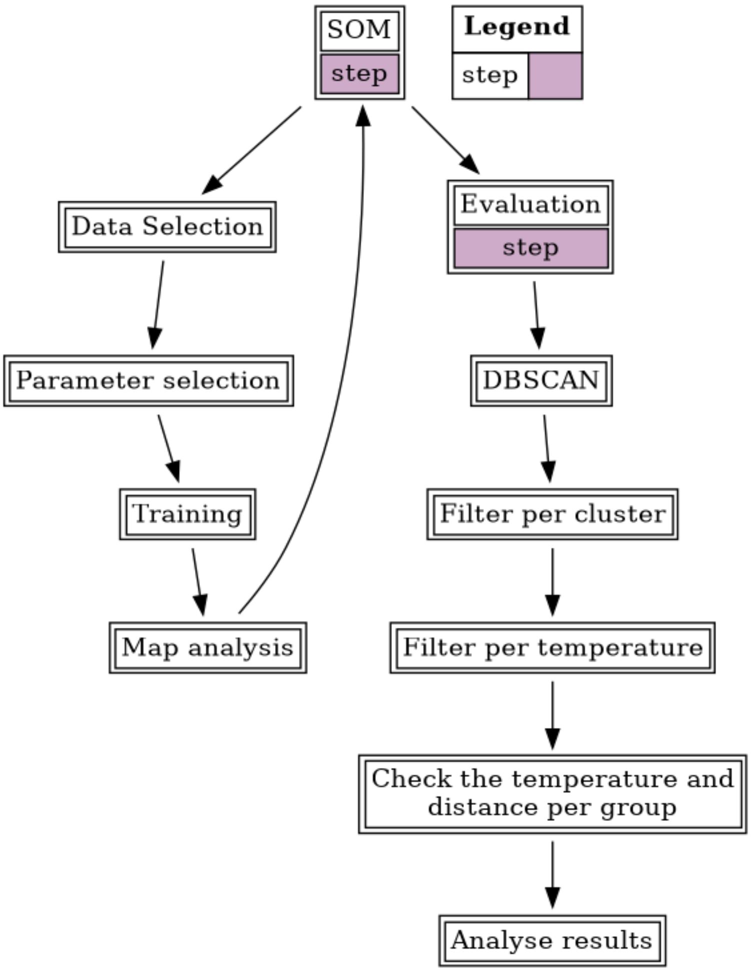
2.3.3 Data selection

From the 15 columns, 5 were chosen. Those being: “timestamp” which included date and hour, “location-lat” which were the latitude of the animal position, “location-long” which were longitude of the animal position, “external-temperature” which were the environment temperature at the moment of the data collected and “tag-local-identifier” which were the identification of the animal.

The external temperature attribute was included because the temperature could be a resource that dominant animals can have more control over the submissive ones. Another attribute that was created was the distance walked for each animal in a given month. This feature was added with the intent of increasing the number of features. The reasoning to add it was that if an animal is in control of a resource, then it would need to walk less.

This data had more than two years of collected data. Since the purpose of this study was to apply to cattle data in a reality closer to the management in Brazilian farms, in this paper, it was chosen to divide the data monthly and apply the SOM method in

Figure 3 – Animal positions



Source: from authors

each month. Another advantage of this approach is that more maps were generated and, therefore, more results could be analyzed.

2.3.4 SOM training

There are many Self-Organizing Maps packages among many programming languages. For this research, it was chosen to use python's package SimpSom (https://github.com/fcomitan
This package was chosen due its simplicity, frequent support and good plotting tools for evaluation.

As many of the SOM packages, in order to create the map for the neural network, the authors had to select some parameters. The most significant one being the size of the map. In this paper, it was chosen to use the rule of thumb proposed in (KOHONEN, 2013) which was to start with 50 rows per node and then adjust more or less nodes according to the results obtained.

The topology chosen was hexagonal for being more illustrative according to Kohonen (2013). The metrics chosen were euclidean because the information intended to be used was the actual distance between each animals and the low dimensionality of the data. Finally, the weights initialization was random.

The parameters chosen were then used to train the SOM and the resulting map was defined satisfactory or not. For this step, a map was considered satisfactory if two conditions were satisfied:

- If there were not too many nodes without a BMU, i.e., an empty node
- If the map was entirely filled with BMU nodes, i.e., if all, or most of the nodes, were filled.

2.3.5 Evaluation

The evaluation was the most important stage because the data used was not labeled. Hence, it was necessary to use the temperature data in order to see if the soft grouping resulted from the SOM was coherent with an dominance hierarchy among the wildebeest.

The data used to train the SOM were from all the animals and divided monthly. After the map was generated, the DBSCAN method for clusterization was used to separate the animals according to the final weights of the SOM. The DBSCAN method was used in order to favor non convex clusters.

After this separation was made, the results expected for a good month, ie, some hierarchy could be found was if three conditions were present:

- The SOM with the DBSCAN could actually find more than one cluster. This was achieved with a map sizes of 13 height and 13 width
- For each cluster, the number of records in it must be at least 10% of the total number of registers in that given month
- From the groups created, the difference from the highest, and lowest temperature must be higher than 1 degree Celsius

The second condition exists in order to ignore noise clusters that are too small. The last condition is intended to enforce clusters with significant difference in the temperatures.

From the resulting months, the next step was to exclude any group in which the temperature of each cluster was not statistically different using a two tailed t-test.

Finally, the last step was to check if the clusters with highest temperature were the ones that moved less in average per animal per day. This criteria was used because it is reasonable to assume that the animal dominating the resource would move less.

2.4 Results

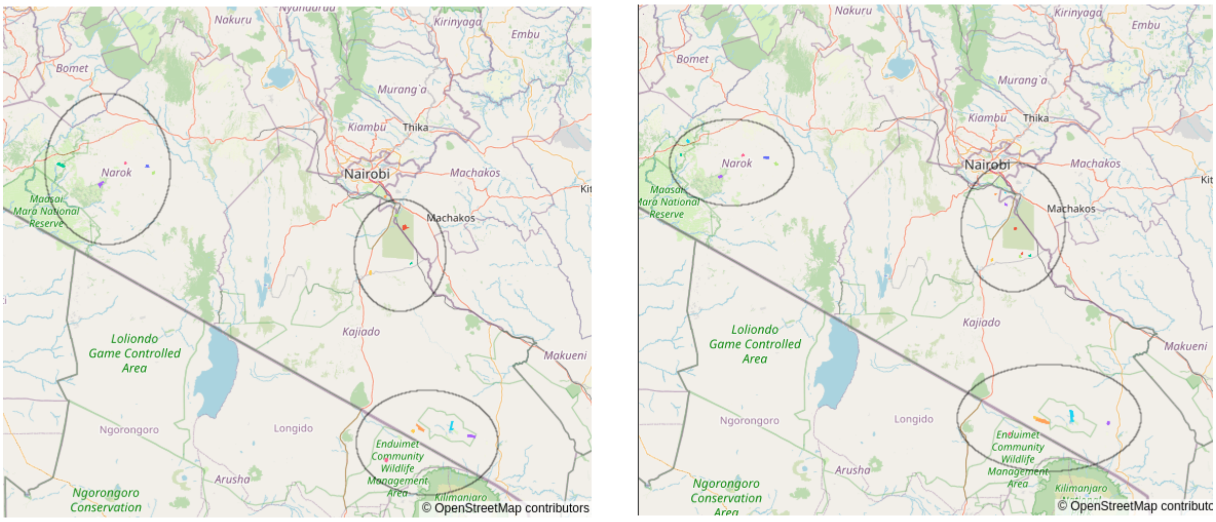
This section contains the results obtained during the research. The results will be presented in 2 sections. First, it will be presented an exploratory analysis and then the results obtained by the methodology proposed with SOM. Finally, an evaluation of the maps generated will be shown. All the code used for this project can be found at https://git.sr.ht/ lucasemmoreira/sominance

2.4.1 Exploratory analysis

The first step was to analyze where the animals were positioned and how close they were to each other. From the 36 animals in the dataset, they grouped in 3 regions as shown in the Figure 4.

Each image illustrates the position pattern of some animals in a given day and each of these animals are represented by a different color.

Figure 4 – Animal positions



Source: from authors

Since there were three groups with animals in relative proximity for existing interaction, it was chosen to apply the proposed methodology to each of the groups separately.

2.4.2 Clusterization results

The methodology proposed in this paper was able to separate the animals in some of the months for each group. In general, for each group, the method was able to create some hierarchy according to the criteria discussed in the methodology in 26.74% among all months and all groups. The Table 2 shows the results.

Table 2 – Overall results from the SOM

Group	Total months	Clustered (number)	Clustered (%)	Hierarchy	Partial Hierarchy	Failed from clustered
1	32	10	31.25%	9.38%	15.63%	20.00%
2	27	2	7.41%	3.70%	3.70%	0.00%
3	27	11	40.74%	14.81%	14.81%	36.36%

Source: Elaborated by the authors

The column “Total months” indicates how many months there are in the simulation for each group. The group 1 has registers from May 2010 up to December 2012 and the groups 2 and 3 start having registers in October 2010. The columns “Clustered” present how many months the SOM was able to successfully separate the animals in different groups. These groups had to attend the two conditions defined in the methodology section (2.3.4). The columns “Hierarchy” and “Partial Hierarchy” shows how many months had some hierarchy found among all the months in the simulation (“Total months”). In other words, from all the months that had available data, how many months the SOM could create some form of hierarchy.

Finally, the last column “Failed from clustered” shows how many months failed to indicate any amount of hierarchy among the months that were clustered (“Clustered (number)”). The intention of this metric is to evaluate how well the methodology proposed could indicate a hierarchy given that the SOM and the DBSCAN could separate the animals.

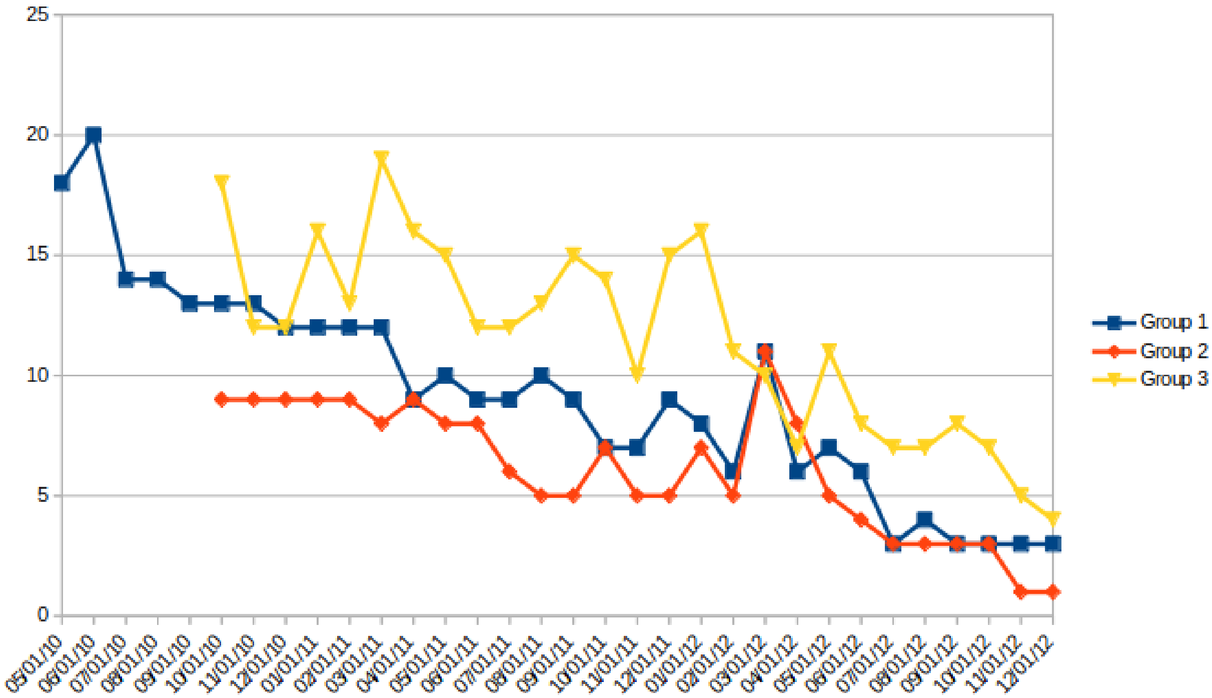
The groups 1 and 3 had similar results. In group 3 there were more months clustered and successfully found hierarchies that were not partial, however from the groups that were clustered by the SOM, group 3 had more failures to obtain some form of hierarchy. Group 2 had a very low rate of clustered months, but one reason to possibly explain that is the lower number of animals in the group in comparison to the others. The Figure 5 shows the evolution of the number of animals during the simulation. Another reason for the group 2 did not have more clustered months was that the difference of the average temperatures of the groups were smaller than 1 degree Celsius.

2.4.3 Map analysis

In this section, a discussion of the maps for each category (successfully hierarchy, partial hierarchy and failed hierarchy) will be presented.

The graphs of the 23 months that were successfully clustered can be extracted and seen in the file. However, for the discussion of this paper only a few graphs will be chosen to be used for the results. To reference each month and group a notation was created to facilitate the understanding. For example, to reference the month of October 2010 of the group 2, the notation used will be g1-2010-10.

Figure 5 – Number of animals



Source: from authors

For each month, four graphs were created. An example can be seen in the Figure 10. The top left graph indicates how each animal records are distributed among the formed clusters. The bottom left graph is the map generated by the SOM. Each color (indicated in the legend by group) represents a record from a given animal. The top right graph is a heat map of the weight difference of the generated map. In general, the dark colors indicate the borders of the clusters to be formed. Finally, the graph on the bottom right shows the map clustered after the DBSCAN method was applied.

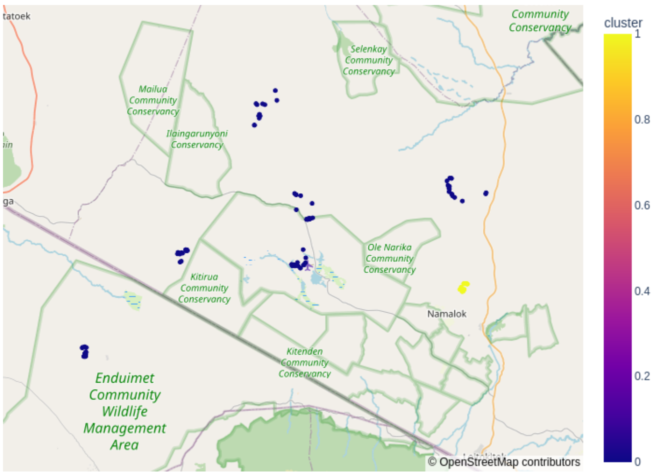
When analyzing the maps, some characteristics were sought in order to be deemed a good result. First, an animal should have most of its records in one cluster. This can be evaluated using the bin graph on the top left. Given one animal, its records on one cluster should be significantly higher than the others. Another possible evaluation can be made using the two maps in the bottom. Ideally, given one animal, i.e. color, on the left map, it should have most of its BMUs in one cluster. In other words, a bad result for an animal would have its BMUs across the whole map belonging to more than one cluster.

The second analysis that can be used to critique the results from the SOM is the number of animals and the number of clusters. If these two numbers are too close, then there is not much information to be extracted from the hierarchy. For instance, if in a given month there are 4 animals and four clusters, then there is little to no information retrieved from the algorithm.

The Figure 9 is an example of a good map. The heat map SOM has a clear boundary and the animals in the different groups have most their records in one cluster.

Another type of image that will be sometimes used in the analysis is a plot of the position of the records of the animals colored by the group they represent. An example is provided in the Figure 6. Each dot represents one record of an animal in that month and the color represents the cluster that the record belongs. The intention of these images is analyze how close each animal is to a body of water which is another resource generally in competition.

Figure 6 – Positioning of the animals from the month g2-2011-02 on February second of 2011



Source: from authors

2.4.3.1 Successful hierarchy

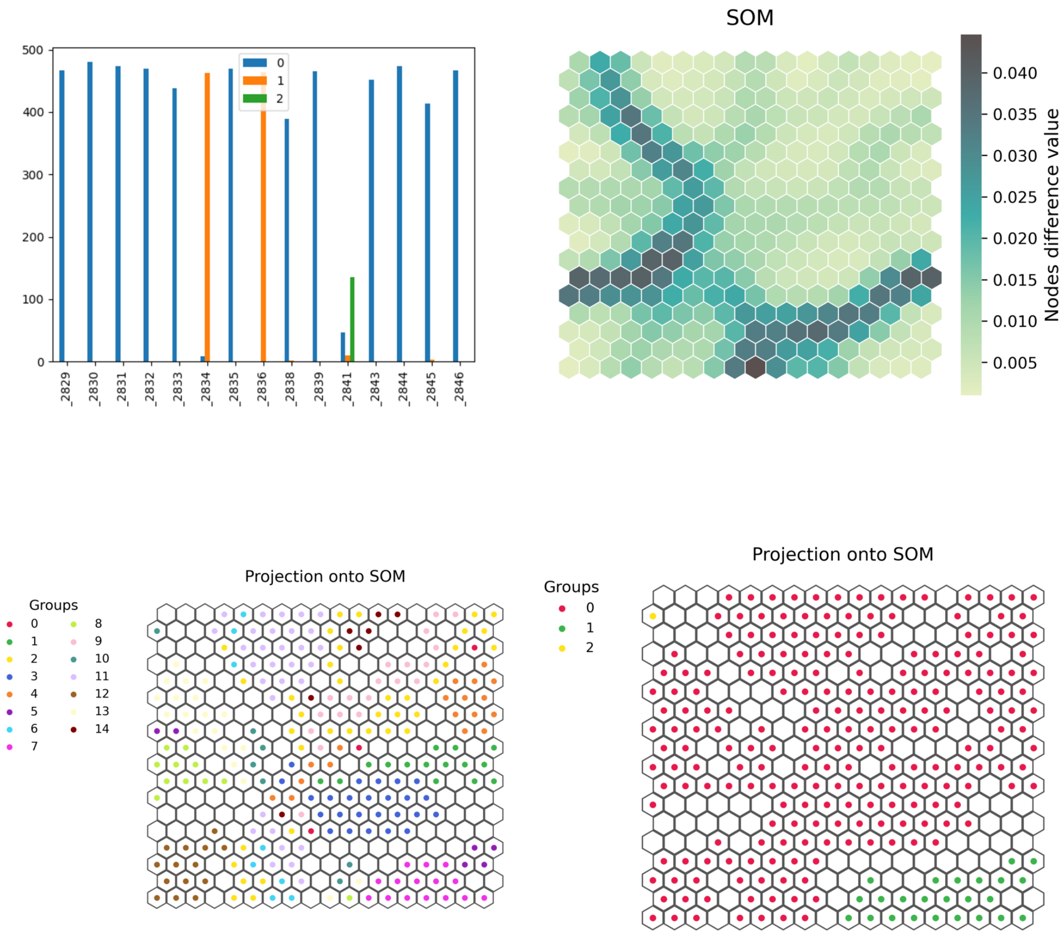
From the four months that were successful in creating an hierarchy, in the month g1-2010-05 there were only several days of records. Therefore, it was hard to understand a pattern from their movement.

The month g1-2010-06 (Figure 7) had a good map and clusterization. With exception of the animal 2841, all animals clearly belonged to one single cluster. The Figure 8 shows that each cluster of animals are close to a body of water.

The month g1-2011-04 (Figure 9) successfully separated one of the animals. Analyzing the movement of the animal 2084 throughout the month, it was possible to see that the animal was actually far away from the herd.

Finally, the last month (Figure 10) in the category had a SOM map that could separate only one animal as well. Although the animals closest to the body of water moved more, there is an airstrip in proximity to that body of water therefore, it can have a lot of noise that made the herd move more.

Figure 7 – Results from the group g1-2010-06



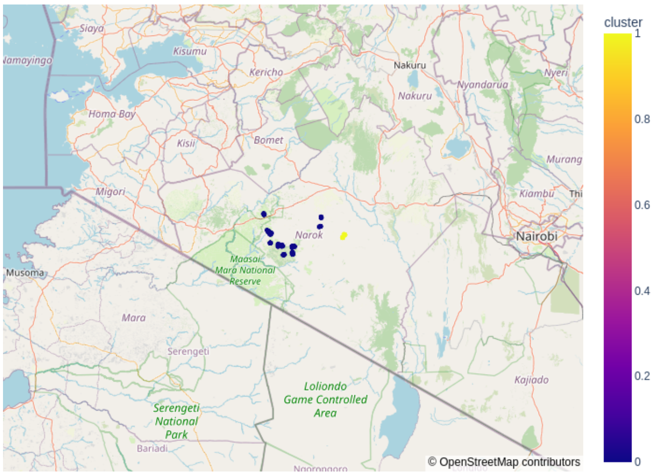
Source: from authors

Overall, in this category, the methodology was able to separate the animals in a coherent manner. The cluster's temperatures were significantly different and the animal positioning were coherent with an hierarchy behavior.

2.4.3.2 Partial Hierarchy

From group1 in the Partial Hierarchy category there were two months that generated a good map - g1-2011-05 and g1-2012-05. In the group 3 of the Partial Hierarchy category, which comprises of the months g3-2010-10 g3-2011-01 g3-2011-03 g3-2011-05 g3-2011-09 g3-2011-12 g3-2012-04, had worse results but happened because those maps had some clusters with very few records. That means that the clusterization method used for these particular months did not perform well to organize the clusters.

Figure 8 – Positioning of the animals from the month g1-2010-06 in one given day of this month



Source: from authors

However, a revisit in these maps to reorganize the clusters following the heat map might produce clusters that still maintains the temperature difference and respect the animals positioning and hierarchy. The Figure 11 illustrates how DBSCAN clustered very small groups that could be seen as part of an adjacent cluster.

Finally, the SOM algorithm in the group 2 of this category (g2-2012-05) successfully separated two of the four animals. However, due to the increasingly missing data of one of those animals might have affected the overall result of this group's month.

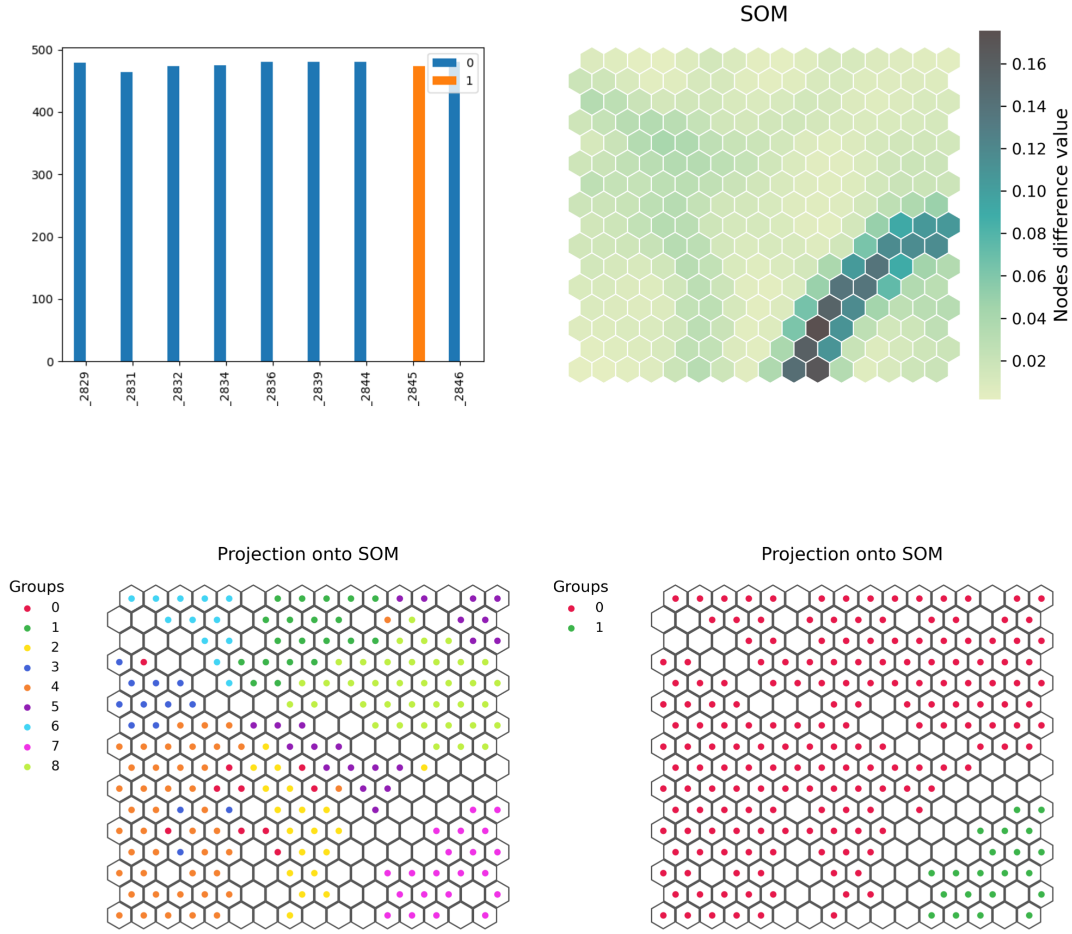
2.4.3.3 Failed to create cluster

The month g1-2011-03 was very similar to the month g1-2011-04 but it had a few days of missing data that impacted the grouping in of the SOM. That might be reason why that month performed poorly in the categorization decided in this research.

The month g1-2012-03 had an clear progression throughout the month. From the movement pattern of the clusters, it was possible to see that in the early stages of the month, both clusters were clumped together and by the end, they were a part showing a possible hierarchy. The Figure 12 shows the positioning of the animals at the beginning of the month (left) and at the end of it (right). It is possible to see how the SOM was able to perceive that the animals drifted apart.

This is an example that would show that this methodology could find an hierarchy that might not be obvious from the start. In addition to that, the map of this month had a boundary in the heat map that would have separated multiple animals which is possible good sign of the map.

Figure 9 – Results from the group g1-2011-04



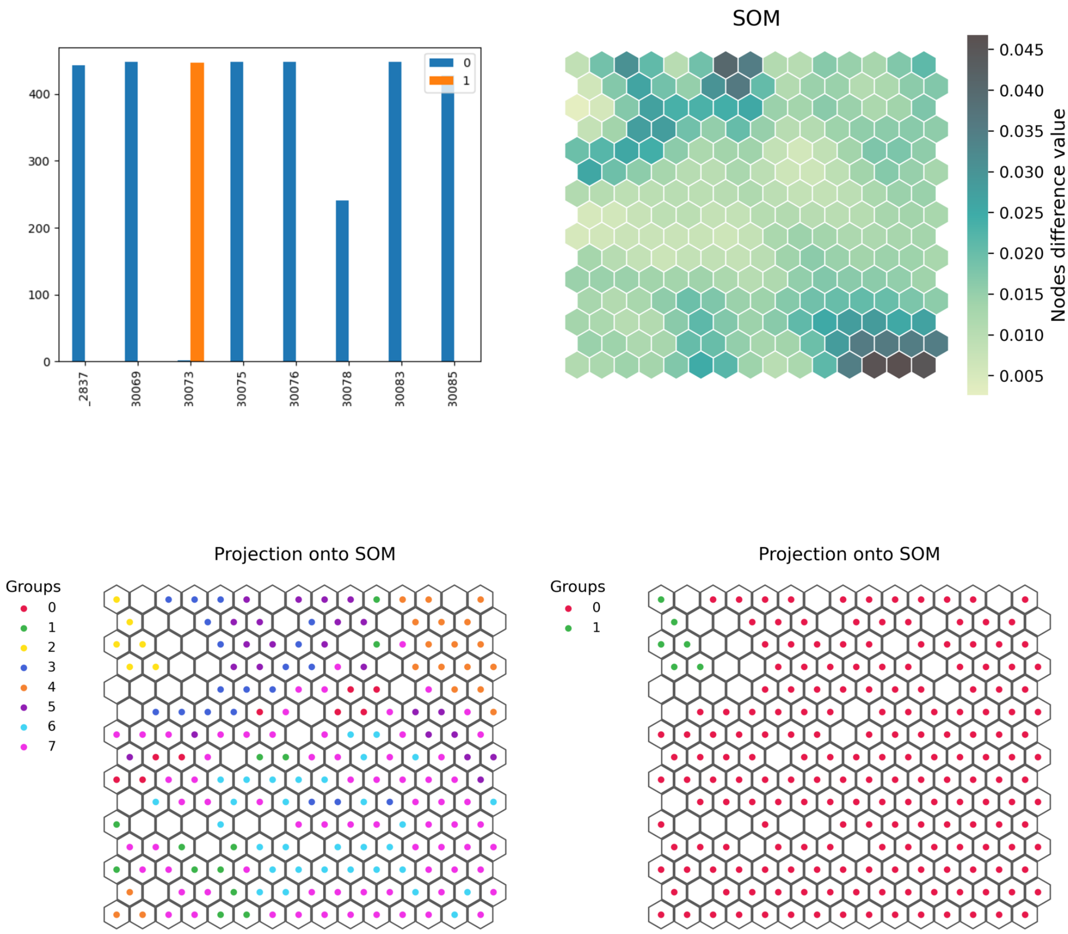
Source: from authors

In Group 3 of this category (g3-2012-01, g3-2012-02, g3-2012-05 and g3-2012-06), with the exception of the group g3-2012-06, had clear distinctions in the SOM that group different animals in different groups. Although, the groups did not follow the assumption that the animals controlling the resource would walk less, it was possible to see that the methodology was capable of differentiating animals that were close in different categories with different average temperatures.

2.5 Final considerations

Overall, the methodology proposed was able to separate the animals in some sort of hierarchy. From the data available, the determination of hierarchy was not precise, however with the provided data in some months this distinction was found.

Figure 10 – Results from the group g2-2011-02



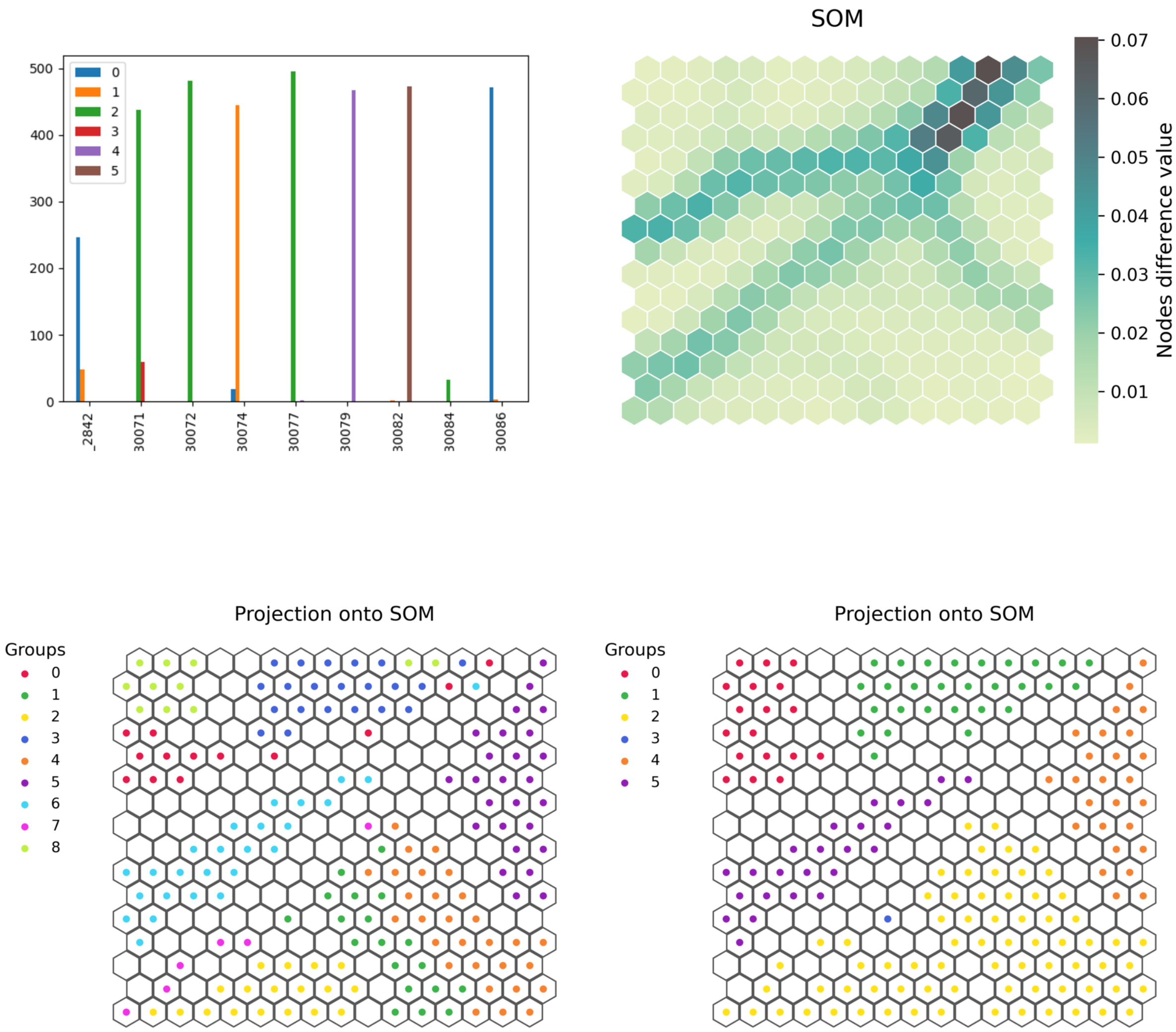
Source: from authors

In the months that did not find an hierarchy, a reinterpretation of the clusters in the map could provide better results. This is a promising analysis to be made in future research.

Other possible routes for this research are

- use more information to feed the SOM. From the data used, the only available information was external temperature and distance walked. Other information could be used, for instance, is weight - the heavier animal the more likely is prevailing over the better food resource. For farm managers another easy information to be gathered is the relative distance of the animal to the water fountain.
- use some technique to handle missing data. From results, it could be seen that there

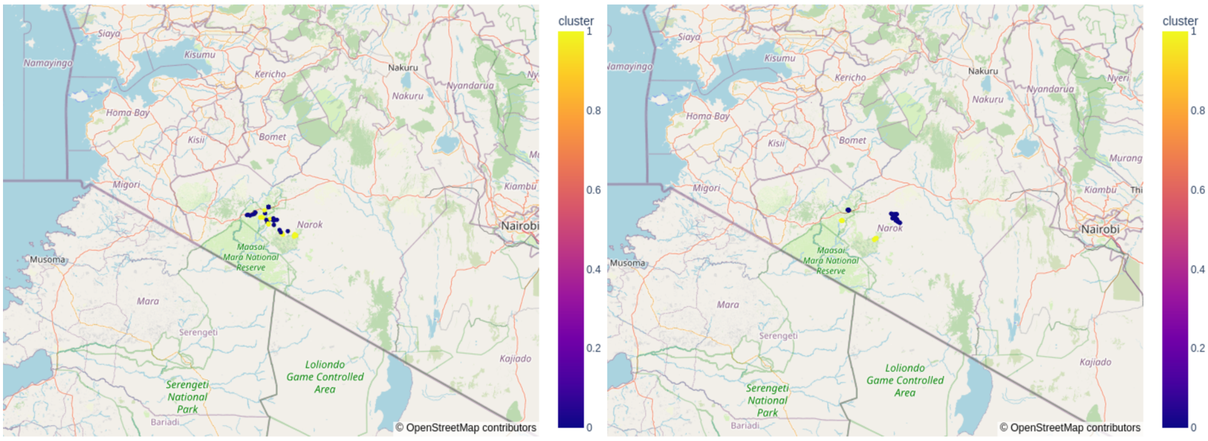
Figure 11 – Results from the group g3-2011-12



Source: from authors

were a few months that had several days with missing information on some animals. This is highly prejudicial for the SOM, therefore some technique to fill this missing data could also improve the results obtained.

Figure 12 – Results from the group g1-2012-03



Source: from authors

3 CONCLUSION

This research proposed to use a “soft” clustering method in order to organize animals in different hierarchies. This could be an valuable information for farm managers to better handle livestock animals in order increase productivity and quality of life for those animals. Because of the time available to dedicate to this project, a few decisions were made in order to be possible to organize the results and evaluate them.

Although there was not enough information about the data (like weight, description of the terrain), still it is safe to assume that the SOM was able to separate the animals in some hierarchy using the geospatial data. In some of the months, the map had a clear boundary and was coherent to the path taken by those animal in the given month.

For future research, it is recommended to have a more detailed data to feed the SOM. Features like average weight of the animal, actual distance from resources like water or food might give more consistent maps to form hierarchies in group of animals. Another take on this research would be to create a different methodology to approach the maps formed by SOM, for example, how to quickly evaluate different groups in one given map.

REFERENCES

ASAN, U.; ERCAN, S. An introduction to self-organizing maps. *In*: **Computational Intelligence Systems in Industrial Engineering: with Recent Theory and Applications**. [*S.l.: s.n.*]: Springer, 2012. p. 295–315.

BAÇÃO, F.; LOBO, V.; PAINHO, M. The self-organizing map, the geo-som, and relevant variants for geosciences. **Computers & geosciences**, Elsevier, v. 31, n. 2, p. 155–163, 2005.

BANKS, E. M. *et al.* Social rank and priority of access to resources in domestic fowl. **Behavioural Processes**, Elsevier, v. 4, n. 3, p. 197–209, 1979.

BOCCIA, M. L.; LAUDENSLAGER, M.; REITE, M. Food distribution, dominance, and aggressive behaviors in bonnet macaques. **American Journal of Primatology**, Wiley Online Library, v. 16, n. 2, p. 123–130, 1988.

CAFAZZO, S. *et al.* Dominance in relation to age, sex, and competitive contexts in a group of free-ranging domestic dogs. **Behavioral Ecology**, Oxford University Press, v. 21, n. 3, p. 443–455, 2010.

DREWS, C. The concept and definition of dominance in animal behaviour. **Behaviour**, Brill, v. 125, n. 3-4, p. 283–313, 1993.

FORIS, B. *et al.* The effects of cow dominance on the use of a mechanical brush. **Scientific Reports**, Nature Publishing Group UK London, v. 11, n. 1, p. 22987, 2021.

HOLEKAMP, K. E.; STRAUSS, E. D. Aggression and dominance: an interdisciplinary overview. **Current Opinion in Behavioral Sciences**, Elsevier, v. 12, p. 44–51, 2016.

HUBBARD, A. J.; FOSTER, M. J.; DAIGLE, C. L. Social dominance in beef cattle—a scoping review. **Applied Animal Behaviour Science**, Elsevier, v. 241, p. 105390, 2021.

KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, IEEE, v. 78, n. 9, p. 1464–1480, 1990.

KOHONEN, T. Essentials of the self-organizing map. **Neural networks**, Elsevier, v. 37, p. 52–65, 2013.

KOUA, E. Using self-organizing maps for information visualization and knowledge discovery in complex geospatial datasets. **Proceedings of 21st international cartographic renaissance (ICC)**, p. 1694–1702, 2003.

PRICE, E. O. **Principles and applications of domestic animal behavior: an introductory text**. [*S.l.: s.n.*]: CABI, 2008.

SHINDE, A.; VERMA, D.; SINGH, N. Social dominance-subordinate relationship in a flock of marwari goats. **Indian Journal of Animal Sciences**, 2004.

SKUPIN, A. Visualizing human movement in attribute space. **Self-Organising Maps: Applications in Geographic Information Science, Agarwal P., Skupin A.,(Eds.). Wiley**, p. 121–135, 2008.

STABACH, J. A. *et al.* Data from: Comparison of movement strategies of three populations of white-bearded wildebeest. 2020.

SYME, G. J.; SYME, L. Social structure in farm animals. **Developments in Animal and Veterinary Sciences (Netherlands). no. 4.**, 1979.

TAKATSUKA, M. An application of the self-organizing map and interactive 3-d visualization to geospatial data. *In*: **Proceedings of the 6th International Conference on GeoComputation**. [*S.l.: s.n.*], 2001. p. 24–26.

VAL-LAILLET, D. *et al.* The concept of social dominance and the social distribution of feeding-related displacements between cows. **Applied Animal Behaviour Science**, Elsevier, v. 111, n. 1-2, p. 158–172, 2008.

WEERD, N. de *et al.* Deriving animal behaviour from high-frequency gps: tracking cows in open and forested habitat. **Plos one**, Public Library of Science San Francisco, CA USA, v. 10, n. 6, p. e0129030, 2015.

ANNEX A – GENERAL RESULTS FROM SOM

cluster	external temperature mean	avg. animal distance per day	number of registers	number of animals	month	year	group	success
3	22.46	0.08	59	2	5	2010	g1	TRUE
4	22.48	0.37	80	2	5	2010	g1	TRUE
0	23.88	0.57	568	10	5	2010	g1	TRUE
6	24.46	0.52	78	1	5	2010	g1	TRUE
2	25.59	0.38	58	1	5	2010	g1	TRUE
1	22.44	1.65	941	5	6	2010	g1	TRUE
0	24.06	3.13	5511	14	6	2010	g1	TRUE
1	23.70	4.09	357	1	3	2011	g1	FALSE
0	25.41	3.96	3501	11	3	2011	g1	FALSE
1	23.98	5.68	474	1	4	2011	g1	TRUE
0	25.40	6.15	3811	8	4	2011	g1	TRUE
1	23.74	6.25	494	1	5	2011	g1	PARTIAL
2	24.63	3.16	496	1	5	2011	g1	PARTIAL
3	24.70	6.64	496	1	5	2011	g1	PARTIAL
0	25.05	4.30	2954	7	5	2011	g1	PARTIAL
4	23.17	6.63	249	1	12	2011	g1	PARTIAL
0	24.34	3.64	937	2	12	2011	g1	PARTIAL
6	24.45	3.47	483	1	12	2011	g1	PARTIAL
5	25.64	3.16	456	1	12	2011	g1	PARTIAL
1	26.43	2.72	486	2	12	2011	g1	PARTIAL
3	26.78	6.26	466	1	12	2011	g1	PARTIAL
0	25.73	4.18	1983	5	1	2012	g1	PARTIAL
1	27.04	3.93	461	1	1	2012	g1	PARTIAL
2	27.34	5.44	493	1	1	2012	g1	PARTIAL
0	25.33	4.43	1867	6	3	2012	g1	FALSE
1	26.79	1.65	1019	5	3	2012	g1	FALSE
2	21.91	0.48	186	1	5	2012	g1	PARTIAL
0	22.29	1.55	397	2	5	2012	g1	PARTIAL
4	22.66	2.71	346	1	5	2012	g1	PARTIAL
3	23.98	1.95	388	1	5	2012	g1	PARTIAL
1	24.84	4.91	764	2	5	2012	g1	PARTIAL
0	21.38	1.26	537	2	6	2012	g1	PARTIAL

cluster	external temperature mean	avg. animal distance per day	number of registers	number of animals	month	year	group	success
1	23.29	2.47	898	3	6	2012	g1	PARTIAL
2	26.91	0.99	169	1	6	2012	g1	PARTIAL
1	26.77	4.06	447	1	2	2011	g2	TRUE
0	28.05	8.11	2905	8	2	2011	g2	TRUE
3	23.68	4.70	243	1	5	2012	g2	PARTIAL
2	25.59	1.65	145	1	5	2012	g2	PARTIAL
1	25.97	4.65	396	1	5	2012	g2	PARTIAL
0	26.05	3.66	655	2	5	2012	g2	PARTIAL
2	24.08	0.96	299	4	10	2010	g3	PARTIAL
3	24.47	1.10	807	4	10	2010	g3	PARTIAL
0	25.04	2.02	1254	5	10	2010	g3	PARTIAL
4	25.74	0.22	161	1	10	2010	g3	PARTIAL
2	24.57	3.29	2255	6	1	2011	g3	PARTIAL
4	24.71	2.47	461	1	1	2011	g3	PARTIAL
5	25.33	2.24	480	1	1	2011	g3	PARTIAL
0	25.60	4.06	1427	4	1	2011	g3	PARTIAL
3	24.55	0.69	179	1	3	2011	g3	PARTIAL
1	24.66	2.30	1760	7	3	2011	g3	PARTIAL
4	25.63	1.10	331	1	3	2011	g3	PARTIAL
0	25.84	2.47	1303	5	3	2011	g3	PARTIAL
6	26.30	0.74	204	1	3	2011	g3	PARTIAL
1	23.34	1.87	459	1	5	2011	g3	PARTIAL
2	23.90	3.39	2500	6	5	2011	g3	PARTIAL
3	24.39	1.92	446	1	5	2011	g3	PARTIAL
0	25.10	4.20	1980	5	5	2011	g3	PARTIAL
0	23.01	2.11	889	6	9	2011	g3	PARTIAL
2	23.28	3.02	1812	5	9	2011	g3	PARTIAL
3	24.66	0.58	322	1	9	2011	g3	PARTIAL
1	24.92	1.55	1102	3	9	2011	g3	PARTIAL
4	24.02	2.67	468	2	12	2011	g3	PARTIAL
1	24.37	1.18	498	4	12	2011	g3	PARTIAL
2	24.54	2.70	1446	4	12	2011	g3	PARTIAL
5	24.74	3.30	473	1	12	2011	g3	PARTIAL
0	25.33	1.21	737	3	12	2011	g3	PARTIAL
0	24.91	2.00	651	4	1	2012	g3	FALSE
3	24.91	1.34	497	4	1	2012	g3	FALSE

cluster	external temperature mean	avg. animal distance per day	number of registers	number of animals	month	year	group	success
2	25.73	2.87	1917	4	1	2012	g3	FALSE
1	26.53	1.55	800	2	1	2012	g3	FALSE
2	24.77	3.95	442	1	2	2012	g3	FALSE
0	25.47	3.17	2209	8	2	2012	g3	FALSE
1	27.05	0.85	266	1	2	2012	g3	FALSE
3	27.55	0.55	355	1	2	2012	g3	FALSE
2	23.20	3.53	409	1	4	2012	g3	PARTIAL
0	24.13	3.80	1369	4	4	2012	g3	PARTIAL
1	24.24	2.50	740	2	4	2012	g3	PARTIAL
0	23.63	2.38	1370	5	5	2012	g3	FALSE
4	23.79	2.27	278	1	5	2012	g3	FALSE
2	24.41	1.71	393	2	5	2012	g3	FALSE
5	25.66	0.40	208	1	5	2012	g3	FALSE
3	21.75	3.87	439	1	6	2012	g3	FALSE
0	22.58	2.41	889	3	6	2012	g3	FALSE
2	22.75	2.14	442	1	6	2012	g3	FALSE
1	22.97	2.72	474	1	6	2012	g3	FALSE
4	23.12	1.64	378	1	6	2012	g3	FALSE

Table 3 – General results