

# On an application of fuzzy logic to building interaction networks between species given geographical data

Víctor Anaya<sup>1</sup>, Víctor Breña<sup>2</sup>, Daniele Colosi<sup>1</sup>, Marisol Flores<sup>1</sup>, Luis Miguel García Velázquez<sup>1</sup>, Adriana Menchaca Méndez<sup>1</sup>, Teresa Patiño Cárdenas<sup>2</sup>, Miguel Raggi<sup>1</sup>, Sergio Rogelio Tinoco Martínez<sup>1</sup>, and César Torres Miranda<sup>1</sup>

<sup>1</sup>Escuela Nacional de Estudios Superiores, Unidad Morelia, UNAM

<sup>2</sup>Centro de Ciencias Matemáticas, UNAM

July 4, 2017

victor\_anaya@enesmorelia.unam.mx, victorb@matmor.unam.mx, dcolosi@enesmorelia.unam.mx, maria.flores@gmail.com, luism\_garcia@enesmorelia.unam.mx, adriana.menchacamendez@gmail.com, lamoradita@gmail.com. mraggi@gmail.com, stinoco@enesmorelia.unam.mx, catomi@gmail.com

No puedo poner los emails!!!!!! No sé cómo! Alguien arréglo antes de que rompa la compu!

## Abstract

We present a novel way of constructing an interaction network for ecological species using tools from fuzzy logic. We present a few (3???) examples: Quercus Oaxaca (?), Loba-whatever, and redes centro, which were obtained by doing bla and bla César?. We also release computational tools that take as input some observational data and output a network built using the methods described in this paper. Furthermore, the code to analyze said network is also freely released. Keywords: fuzzy logic, ecology, networks

Mathematics Subject Classification: ???

## 1 Introduction

In [PJ94] and [TM14] the authors do bla bla bla...

In this paper, we improve on that because of bla bla bla, and apply the results to bla bla and bla.

All the source code used for this paper can be freely downloaded and used (under the terms of the license) from

<https://www.github.com/mraggi/FuzzyLogicEcology>

Traditionally, the way to build an interaction network is to use circles or polygons or whatever, but that has the problems of... César

## 1.1 A brief introduction to the tools of fuzzy logic

In this section we introduce the reader to the fuzzy logic tools we used to build the interaction network. For a more complete treatment on the topic, see [?] or [?].

We differentiate between regular sets (*crisp* sets) and *fuzzy* sets. A fuzzy set is a set in which membership is not a boolean value, but rather is a number between 0 and 1, where a membership of 0 means that the element is not in the set and 1 means the element is definitely in the set. Values in between indicate how much an element belongs to a set.

Formally,

**Definition 1.1.** A **fuzzy set**  $A = (X, \mu_A)$  is a set  $X$  together with a function  $\mu_A : X \rightarrow [0, 1]$ . In this case we call  $X$  the *ground set* and  $\mu_A$  the *membership function* of  $A$ . Sometimes we refer to  $\mu_A$  as the fuzzy set.

We wish to define the standard set operations ( $\subset, \setminus, \cap, \cup$ ) on fuzzy sets.

**Definition 1.2.** Let  $A$  and  $B$  be fuzzy sets on the same ground set  $X$ . Then,

- $A \subset B$  iff  $\mu_A(x) \leq \mu_B(x)$  for every  $x \in X$ .
- $A \setminus B$  is defined by the characteristic function  $\mu_{A \setminus B}(x) = \max(0, \mu_A(x) - \mu_B(x))$ .

The above definitions are the only standard definition of  $\subset$  and  $\setminus$ , but the definition of  $\cap$  and therefore  $\cup$  varies, as there are many ways to do so in a consistent and logical manner. Usually one gives the definition of  $\cap$  as  $\mu_{A \cap B} = \mathbb{I}(a, b)$ , where  $\mathbb{I} : \mathbb{R}^2 \rightarrow [0, 1]$  is a function that satisfies the same properties as intersection when evaluated in  $\{0, 1\}$ . The definition of  $\cup$  follows. paperize!.

**Definition 1.3.** A *consistent* function  $\mathbb{I} : [0, 1]^2 \rightarrow [0, 1]$  is one that

1.  $\mathbb{I}(a, b) = \mathbb{I}(b, a)$  for all  $a, b$ . (symmetry)
2.  $\mathbb{I}(0, b) = 0$  for all  $b$ .
3.  $\mathbb{I}(1, b) = b$  for all  $a$ .
4.  $\mathbb{I}(a, b) \leq \mathbb{I}(c, b)$  if  $a \leq c$ .

Once the intersection is defined as  $\mu_{A \cap B}(p) := \mathbb{I}(\mu_A(p), \mu_B(p))$ , the definition of  $\cup$  easily follows as in the inclusion-exclusion principle:  $\mu_{A \cup B} = \mu_A + \mu_B - \mu_{A \cap B}$ .

It can be easily checked that if one defines intersection this way with a consistent  $\mathbb{I}$ , it satisfies many of the common properties of sets, such as commutativity for  $\cap$  and  $\cup$ , associativity, De Morgan's laws, etc. when the corresponding crisp set properties are satisfied as well paperizar!.

In this work we will consider two different definitions of  $\mathbb{I}$ . Both have advantages and disadvantages.

The first definition we shall consider involves defining the operations akin to when considering independent sets in probability:  $\mathbb{I}(a, b) = ab$ .

**Definition 1.4.** •  $A \cap B$  is defined by the membership function  $\mu_{A \cap B}(x) := \mu_A(x)\mu_B(x)$

•  $A \cup B$  is defined by the membership function  $\mu_{A \cup B}(x) := \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x)$

The second definition, also quite common, is to define  $\mathbb{I}(a, b) = \min(a, b)$ . In this scenario, we observe that  $\mu_{A \cup B}(p) = \max(\mu_A(p), \mu_B(p))$ .

## 2 Building the network

The main idea behind the method is that instead of considering the distribution area of a species as a crisp set, we consider the distribution as a fuzzy set. The motivation behind this is that given an observation of an individual of a certain species, the area of effect of this individual in its interaction with other species will decay the further away one moves from this observation. **Este enunciado está terrible. Re-escribir! Y pasar para antes, probablemente.**

Let  $\Omega \subset \mathbb{R}^2$  denote the full area of distribution we are considering.

Let  $P$  denote a point in  $\Omega$  and  $C > 0$  be real constant number. We define  $f_P : \Omega \rightarrow [0, 1]$  in the following way:

$$f_P(Q) = e^{-C[d(P, Q)^2]}.$$

where  $d(P, Q)$  denotes the (Euclidean) distance from  $P$  to  $Q$ .

Now, for a certain species  $\mathcal{A}$ , let  $A \subset \Omega$  denote the set of observed locations of an individual of that species. It is common to find such data when doing field work of this type and ... **César? O quizás en la intro y aquí nada más decir “As we previously mentioned, ...”.**

**Definition 2.1.** We define the **area of distribution** of species  $\mathcal{A}$  as the fuzzy set given by the following membership function:

$$\mu_A = \bigcup_{a \in A} f_a$$

Following the standard way of constructing interaction networks based on geographical data ([TM14], [?]), we build a weighted directed graph  $D = (V, E)$  where  $V$  is the set of species, and there is an edge  $\mathcal{A} \rightarrow \mathcal{B}$  if their respective areas of distribution of  $\mathcal{A}$  and  $\mathcal{B}$  intersect, and the weight of this edge is defined as

$$w(\mathcal{A} \rightarrow \mathcal{B}) = \frac{\text{Area}(\mathcal{A} \cap \mathcal{B})}{\text{Area}(\mathcal{B})}$$

Translating the above definition to the fuzzy world, we may calculate the weight as follows: if  $\mu_A$  and  $\mu_B$  are the respective membership functions of  $\mathcal{A}$  and  $\mathcal{B}$  as defined above, when  $\mathbb{I}$  is the product function,

$$w_P(\mathcal{A} \rightarrow \mathcal{B}) := \left( \int_{\Omega} \mu_A \mu_B \right) / \left( \int_{\Omega} \mu_A \right)$$



Figure 2.1: Visualization of the area of distribution of a species

and when  $\mathbb{I}$  is the min function,

$$w_M(\mathcal{A} \rightarrow \mathcal{B}) := \left( \int_{\Omega} \min(\mu_A, \mu_B) \right) / \left( \int_{\Omega} \mu_A \right)$$

## 2.1 Calculating weights numerically

We are left with the following computational problem: Given a set of species and for each species a set of points representing the observational data of the species in question, calculate the weighted digraph as above.

Obviously an analytical approach is a terrible way to do it **Bah, justifiquen esto uds. Daniele? Alguien?**

So we rely on numerical methods instead. First we'll describe how to find  $\mu_A$  for each species  $\mathcal{A}$ , and then we proceed to find  $w(\mathcal{A} \rightarrow \mathcal{B})$  for each pair of species  $\mathcal{A}, \mathcal{B}$ .

Divide  $\Omega$  into an  $n \times n$  grid. Then, each  $\mu_A$  can be discretized and represented by an  $n \times n$  matrix of real numbers. The higher  $n$  is, the better the approximation. For all our experimental results, we found that  $n \approx 5000$  is more than enough to get an accurate approximation. **Alguien se quisiera aventar el análisis de error? No? That's what I thought.**

## 2.2 Normalizing

In a brute force approach, since for every point in the grid we have to consider every observation of said species, calculating the matrix associated with  $\mu_A$  takes at least  $\Theta(|A|n^2)$  time. To speed up computations by sacrificing **(a tiny amount of)** accuracy, note that

$$\mu_A \leq \sum_{a \in A} f_a.$$

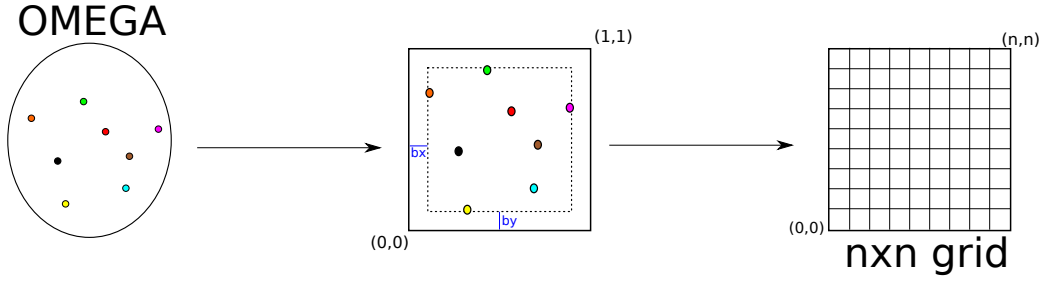
We note that the exponential decay in the formula for  $f_a$  gives a quick way to only affect the area near point  $a$ . Therefore, if we admit an error of  $\varepsilon > 0$  for each point in the grid, we

only need to compute  $f_a(p)$  for points  $p$  near  $a$  inside a circle of radius  $R$ , where

$$R := f_a^{-1}(\varepsilon/|A|) = \sqrt{-\log\left(\frac{\varepsilon}{|A|C}\right)}$$

In this way, calculating the matrix associated with  $\mu_A$  takes  $\Theta(|A|R^2 + n^2)$ , a considerable speedup. This assumes, of course, that each point in the grid can be calculated in  $\Theta(|A|)$  time, which is not immediate when  $\mathbb{I}$  is the product. In the next section we make it so. **paperize.**

With this in mind, we proceed to normalize as follows:



First, let  $O = (mx, my)$  and  $W = (Mx, My)$ , where  $mx, my, Mx, My$  denote the minimum and maximum points in the dataset. Let  $F = W - O$ .

Then,  $b_x = R/|F_x|$  and  $b_y = R/|F_y|$  are the  $x$  and  $y$  borders respectively. Any element of the grid outside the border would, by the definition of  $R$ , contain a number that is less than  $\varepsilon$  and therefore can be ignored.

Also,  $C$  has to be transformed into  $C_x$  and  $C_y$  by the equation **look it up in the code**, because of **Daniele said so. I agreed he was correct, but I don't remember.**

## 2.3 Finding $\mu_A$

Finding the matrix representing  $\mu_A$  is straightforward when considering  $\mathbb{I}$  as the min function. For the rest of this section, we consider only the case when  $\mathbb{I}$  is the product function.

When  $\mathbb{I}$  is the product function, however, things get a little more tricky **paperize.** By the inclusion-exclusion principle we have

$$\mu_A = \sum_{a \in A} f_a - \sum_{a, b \in \binom{A}{2}} f_a f_b + \sum_{a, b, c \in \binom{A}{3}} f_a f_b f_c + \cdots + (-1)^{N+1} \prod_{a \in A} f_a$$

We factorize  $\mu_A$  as follows:

$$\mu_A = 1 - \prod_{a \in A} (1 - f_a) \tag{2.1}$$

## 2.4 Finding $w$

Once we have  $\mu$  for every species, it's time to calculate  $\int \mu_A \mu_B$  for every pair of species  $\mathcal{A}$ ,  $\mathcal{B}$ . The straightforward approach has a complexity of  $O(s^2 n^2)$ , where  $s$  is the number of species and  $n$  is the size of the grid.

We may speed up this computation when  $\mathbb{I}$  is the product with the following trick.

Note that

$$\int \mu_A \mu_B \approx \sum_{0 \leq x < n} \sum_{0 \leq y < n} \mu_A[x, y] \mu_B[x, y]$$

If we flatten the matrix  $\mu_A$  and  $\mu_B$  to vectors  $\nu_A$  and  $\nu_B$ , this last sum is simply the dot product  $\nu_A \cdot \nu_B$ .

So consider the  $s \times n^2$  matrix

$$M = \begin{bmatrix} \leftarrow & \nu_0 & \rightarrow \\ \leftarrow & \nu_1 & \rightarrow \\ & \dots & \\ \leftarrow & \nu_{s-1} & \rightarrow \end{bmatrix}.$$

Then,  $MM^T$  is an  $s \times s$  matrix whose  $(i, j)$ -th entry contains  $\int \mu_A \mu_B$ . This can be done efficiently with fast matrix multiplication algorithms ([?]) (Marisol, tú te los sabes seguramente, pon unas palabritas de ellos y di que son super chidos y que no se que). Unfortunately, when  $\mathbb{I} = \min$  this speedup is not possible, as fast matrix multiplication relies on the fact that  $+$  and  $\times$  are distributive.

Note also that the algorithms described here are highly parallelizable, since calculating the matrix associated with  $\mu_A$  for one species does not rely on observations from any other species. Indeed, the provided software benefits considerably by using a computer with many cores. **paperize?**

## 3 Experiments and results

### 3.1 Digraphs

The following digraphs were obtained as a result of the algorithms described in the previous section. (mm... damos las gráficas? La más chica es una matriz de  $56 \times 56$ , aunque podríamos dar sólo las aristas más pesadas... no sé)

### 3.2 Analysis of the digraphs

This graphs are cool because here is the robustness and the analysis revealed deep and important truths (Yo les doy los resultados del análisis. Uds. escriben por qué, para qué, cómo, cuándo, etc.)

## 4 Conclusions

Finally, we conclude this awesome work. Please accept this paper!!

## Acknowledgements

This project was supported by PAPIIT IA106316.

## References

- [PJ94] Fulano Perez and Sutano Juarez, *Fuzzy logic and bla bla bla*, Ecology today! **10** (1994).
- [TM14] César Andrés Torres Miranda, *Importancia de los ensambles de especies y de los procesos geológico-climáticos en la comprensión de los mecanismos de evolución de la sección lobatae del género quercus en américa del norte*, Ph.D. thesis, Instituto de Investigaciones en Ecología, Universidad Nacional Autónoma de México, 2014.

## A On the mathematical expression of the area distribution

We consider a species  $A$  and denote by  $N$  the number of its individuals; then the expression (2.1) reads:

$$\mu_A = \sum_{i=1}^N f_i - \sum_{i \neq j} f_i f_j + \sum_{i \neq j \neq k, i \neq k} f_i f_j f_k + \cdots + (-1)^{N-1} \prod_{i=1}^N f_i, \quad (\text{A.1})$$

or, in more compact notation,

$$\mu_A = \sum_{n=1}^N (-1)^{n+1} \sum_{1 \leq i_1 < \cdots < i_n} f_{i_1} \cdots f_{i_n}. \quad (\text{A.2})$$

Now we compute the integral of  $\mu_A$  over all of space, i.e. over  $\mathbb{R}^2$ :

$$\begin{aligned}
& \int_{\mathbb{R}^2} dx dy \mu_A(x, y) \\
&= \sum_{i=1}^N \int_{\mathbb{R}^2} dx dy f_i(x, y) - \sum_{i \neq j} \int_{\mathbb{R}^2} dx dy f_i(x, y) f_j(x, y) \\
&+ \sum_{i \neq j \neq k, i \neq k} \int_{\mathbb{R}^2} dx dy f_i(x, y) f_j(x, y) f_k(x, y) + \cdots + (-1)^{N-1} \int_{\mathbb{R}^2} dx dy \prod_{i=1}^N f_i(x, y), \\
&= \sum_{n=1}^N (-1)^{n+1} \sum_{1=i_1 < \cdots < i_n} \int_{\mathbb{R}^2} dx dy f_{i_1}(x, y) \cdots f_{i_n}(x, y). \tag{A.3}
\end{aligned}$$

Notice that if the  $f_i$  are Gaussian distributions, then the first integral in the r.h.s. of the second line of the above expression results to be

$$\int_{\mathbb{R}^2} dx dy f_i(x, y) = \frac{\pi}{C}. \tag{A.4}$$

We consider the integral of the product of  $n$  Gaussians:

$$\int_{\mathbb{R}^2} dx dy f_{i_1}(x, y) \cdots f_{i_n}(x, y) = \int_{\mathbb{R}^2} dx dy \exp \left[ -C \sum_{\alpha=1}^n ((x - x_{i_\alpha})^2 + (y - y_{i_\alpha})^2) \right]. \tag{A.5}$$

Notice that

$$\sum_{\alpha=1}^n (x - x_{i_\alpha})^2 = nx^2 - 2x \underbrace{\sum_{\alpha=1}^n x_{i_\alpha}}_{X_{1,n}} + \underbrace{\sum_{\alpha=1}^n x_{i_\alpha}^2}_{X_{2,n}}. \tag{A.6}$$

So, using the obvious notation,

$$\begin{aligned}
& \int_{\mathbb{R}^2} dx dy f_{i_1}(x, y) \cdots f_{i_n}(x, y) \\
&= \int_{\mathbb{R}^2} dx dy \exp \left[ -Cn(x^2 + y^2) + 2C(X_{1,n}x + Y_{1,n}y) - C(X_{2,n} + Y_{2,n}) \right], \\
&= \frac{\pi}{nC} \exp \left[ \frac{C(X_{1,n}^2 + Y_{1,n}^2)}{n} - C(X_{2,n} + Y_{2,n}) \right], \\
&= \frac{\pi}{nC} \exp \left[ \frac{C((\sum_{\alpha=1}^n x_{i_\alpha})^2 + (\sum_{\alpha=1}^n y_{i_\alpha})^2)}{n} - C \sum_{\alpha=1}^n (x_{i_\alpha}^2 + y_{i_\alpha}^2) \right], \\
&= \frac{\pi}{nC} \exp \left[ -\frac{C}{n} \sum_{1=\alpha < \beta}^n ((x_{i_\alpha} - x_{i_\beta})^2 + (y_{i_\alpha} - y_{i_\beta})^2) \right], \\
&= \frac{\pi}{nC} \exp \left[ -\frac{C}{n} \sum_{1=\alpha < \beta}^n d_{i_\alpha i_\beta}^2 \right], \tag{A.7}
\end{aligned}$$



where  $d_{i_\alpha i_\beta}$  is the (euclidean) distance between point  $P_{i_\alpha} = (x_{i_\alpha}, y_{i_\alpha})$  and point  $P_{i_\beta} = (x_{i_\beta}, y_{i_\beta})$ . Finally

$$\int_{\mathbb{R}^2} dx dy \mu_A = \sum_{n=1}^N (-1)^{n+1} \sum_{1=i_1 < \dots < i_n}^N \frac{\pi}{nC} \exp \left[ -\frac{C}{n} \sum_{1=\alpha < \beta}^n d_{i_\alpha i_\beta}^2 \right]. \quad (\text{A.8})$$

Analogously we obtain the integral of the product of two membership functions

$$\begin{aligned} \int_{\mathbb{R}^2} dx dy \mu_A \mu_B &= \sum_{n=1}^{N_A} \sum_{\tilde{n}=1}^{N_B} (-1)^{n+\tilde{n}} \sum_{1=i_1 < \dots < i_n}^{N_A} \sum_{1=\tilde{i}_1 < \dots < \tilde{i}_{\tilde{n}}}^{N_B} \frac{\pi}{(n+\tilde{n})C} \\ &\exp \left[ -\frac{C}{n+\tilde{n}} \left( \sum_{1=\alpha < \beta}^n d_{i_\alpha i_\beta}^2 + \sum_{1=\alpha < \beta}^{\tilde{n}} d_{\tilde{i}_\alpha \tilde{i}_\beta}^2 + \sum_{\alpha=1}^n \sum_{\tilde{\alpha}=1}^{\tilde{n}} d_{i_\alpha \tilde{i}_{\tilde{\alpha}}}^2 \right) \right]. \end{aligned} \quad (\text{A.9})$$