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

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

[1]Escuela Nacional de Estudios Superiores, Unidad Morelia, UNAM
[2]Centro de Ciencias Matemáticas, UNAM

July 9, 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 arréglelo 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, and a web service is launched, where a user may submit a text file with data and get a network back. 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

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

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 $\mu_A : X \to [0, 1]$. In this case we call X the *ground set* and μ_A the *membership function* of A. Sometimes we refer to μ_A as the fuzzy set.

We wish to define the standard set operations $(\subset, \backslash, \cap, \cup)$ 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 $x \in X$.

- $A \backslash B$ is defined by the characteristic function $\mu_{A \backslash B}(x) = \max(0, \mu_A(x) - \mu_B(x))$.

The above definitions are the only standard definition of $\subset$ and $\backslash$, 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 $\mu_{A \cap B}(p) := \mathbb{I}(\mu_A(p), \mu_B(p))$, where $\mathbb{I} : [0, 1]^2 \to [0, 1]$ is a function that satisfies the same properties as intersection when evaluated in $\{0, 1\}$. The definition of $\cup$ follows. paperize! Luis Miguel?.

Definition 1.3. A *consistent* function $\mathbb{I} : [0, 1]^2 \to [0, 1]$ is one that for all $a, b, c \in [0, 1]$

1. $\mathbb{I}(a, b) = \mathbb{I}(b, a)$. (symmetry)

2. $\mathbb{I}(0, b) = 0$.

3. $\mathbb{I}(1, b) = b$.

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 paperize! Luis Miguel?.

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

The first definition we shall consider involves defining the operations akin to when considering independent sets in probability: $\mathbb{I}(a, b) = ab$, so that $A \cap B$ is defined by the membership function $\mu_{A \cap B}(x) := \mu_A(x)\mu_B(x)$, and therefore $A \cup B$ 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. This sentence is terrible. Rewrite. Luis Miguel

Let $\Omega \subset \mathbb{R}^2$ denote the full area of distribution we are considering. Let P denote a point in Ω and $C > 0$ be real constant number. We define $f_P : \Omega \to [0, 1]$ in the following way:

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

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:

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

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 $\mathcal{A} \to \mathcal{B}$ if their respective areas of distribution of $\mathcal{A}$ and $\mathcal{B}$ intersect, and the weight of this edge is defined as

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

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

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

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

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

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

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 μ_A for each species $\mathcal{A}$, and then we proceed to find $w(\mathcal{A} \to \mathcal{B})$ for each pair of species $\mathcal{A},\mathcal{B}$.

Divide Ω into an $n \times n$ grid. Then, each μ_A can be discretized and represented by an $n \times n$ matrix of real numbers, where every point in the $n \times n$ grid contains a number in $[0,1]$. 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, in the sense that there is no noticeable difference between the edge weights when $n = 5000$ and when $n = 20,000$. Alguien se quisiera aventar un análisis de error? No? That's what I thought.

2.2 Normalizing

Before going into detail of how to normalize the data in order to calculate the integrals numerically, we discuss allowing some tiny error into our calculations in order to reduce Ω to a bounded set. This discussion will have the added benefit of making the computation much faster. paperize. And probably re-write. Daniele!! I tried, but couldn't quite get it.

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 μ_A takes at least $\Theta(|A|n^2)$

time. To speed up computations by sacrificing (a tiny amount of) accuracy, note that

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

Uh... paperize the "a tiny amount of" part. Daniele

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 μ_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, but this is not immediately obvious how to do so when $\mathbb{I}$ is the product. In the next section we make it so. paperize. Marisol o Adriana o Sergio.

With this in mind, we proceed to normalize as follows:
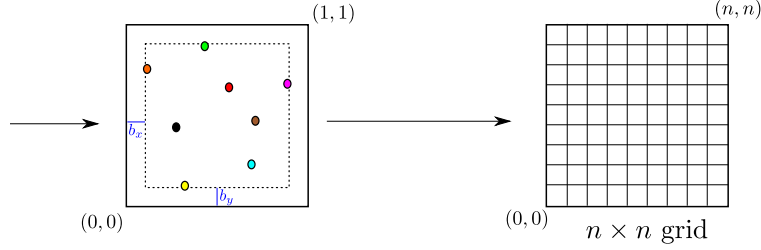
The second function just involves multiplying each coordinate by n so it is easy. paperize. Daniele

Consider the bounding box in Ω of the points in the data set. That is, let $O = (mx, my)$ and $W = (Mx, My)$, where mx, my, Mx, My denote the minimum and maximum x and y values in the dataset respectively (of all observations for all species in question) respectively.

Let $F = W - O$. Then, $b_x = R/|F_x|$ and $b_y = R/|F_y|$ are the x and y borders respectively in $[0, 1] \times [0, 1]$. For any point falling outside the border, the value of μ would, by the definition of R, be a number that is less than ε and therefore can be ignored. Every point in the dataset gets normalized in this way.

Also, C has to be transformed into $C_x := C\frac{F_x^2}{n^2}$ and $C_y := C\frac{F_y^2}{n^2}$, because of Daniele said so. I agreed he was correct, but I don't remember.

6

2.3 Finding μ_A

Computing the matrix representing μ_A is straightforward when considering $\mathbb{I}$ as the min function: For each point in the grid we must simply find the closest point $a \in A$ and calculate f_a.

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

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

We factor μ_A as follows:

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

2.4 Finding w

Once we have μ 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

$$n^2 \int \mu_A \mu_B \approx \sum_{0 \le x < n} \sum_{0 \le y < n} \mu_A[x,y] \mu_B[x,y]$$

Let ν_A and ν_B denote vectors that result from flattening μ_A and μ_B. Then the r.h.s. of the previous equation is simply the dot product $\nu_A \cdot \nu_B$.

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

$$M = \begin{bmatrix} \longleftarrow & \nu_0 & \longrightarrow \\ \longleftarrow & \nu_1 & \longrightarrow \\ & \cdots & \\ \longleftarrow & \nu_{s-1} & \longrightarrow \end{bmatrix}.$$

Then, MM^T is an $s \times s$ matrix whose (i,j)-th entry contains $\nu_A \cdot \nu_B \approx n^2 \int \mu_A \mu_B$. However, MM^T can be computed 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 μ_A for one species does not rely on observations from any other species, and the fast matrix multiplication algorithms are highly paralellizable paperize. Marisol: do this. Indeed, the provided software benefits considerably by using a computer with many cores. paperize?

2.5 Advantages and disadvantages of min over product

We gave two parallel ways to construct the distribution network: using products and using min. In this section we discuss the advantages and disadvantages of using each.

The clearest disadvantage of using the product is that $A \cap A \neq A$. So the weighted edge $w_P(\mathcal{A} \to \mathcal{A}) \neq 1$. This sucks. In general, in our experimental data, the edge weights become quite small. A way to alleviate this is to consider

$$\widehat{w_P}(\mathcal{A} \to \mathcal{B}) := \left(\int_\Omega \mu_A \mu_B \right) / \left(\int_\Omega \mu_A^2 \right)$$

This gives much better edge weights paperize... justify? Luis Miguel?, but has the disadvantage that an edge weight may be greater than 1. Using our experimental data. Since the edge weight $w(\mathcal{A} \to \mathcal{B})$ is supposed to represent how what percentage of the area of distribution of $\mathcal{A}$ is contained in the area of distribution of $\mathcal{B}$, a value greater than 1 would mean that $\mathcal{A}$ is more strongly contained in $\mathcal{B}$ than in itself. This makes sense!! It's not bad!

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 56×56, aunque podríamos dar sólo las aristas más pesadas... no sé)

3.2 Analysis of the digraphs

Is this section for second paper? Or what?

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!! Later

Acknowledgements

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

Me gustaría decir aquí que usamos estas expresiones como unit tests del programa. Adriana? Marisol? Sergio?

In this appendix we provide an analytic expression of the integral of the (product of the) membership functions in an ideal case, namely where the integral is extended over $\mathbb{R}^2$. Although this is unrealistic because the species are distributed on the surface of the Earth, we consider such an idealisation justified by the exponential decay of the functions f_P defined above.

Suggestion: so far, we have been using f_a where a is a point, not f_i where $i \in \mathbb{Z}$.

Given a species A, we denote by N_A the number of its individuals; then the expression (**??**) reads:

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

or, in more compact notation,

$$\mu_A = \sum_{n=1}^{N_A} (-1)^{n+1} \sum_{1 = i_1 < \cdots < i_n}^{N_A} f_{i_1} \cdots f_{i_n}. \tag{A.2}$$

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

$$\int_{\mathbb{R}^2} \mathrm{d}x \, \mathrm{d}y \, \mu_A(x, y) = \sum_{n=1}^{N_A} (-1)^{n+1} \sum_{1 = i_1 < \cdots < i_n}^{N_A} \int_{\mathbb{R}^2} \mathrm{d}x \, \mathrm{d}y \, f_{i_1}(x, y) \cdots f_{i_n}(x, y). \tag{A.3}$$

Notice that 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

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

We consider the integral of the product of n Gaussian distributions:

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

Notice that

$$\sum_{\alpha=1}^{n}\left((x-x_{i_\alpha})^2 + (y-y_{i_\alpha})^2\right) = n(x^2+y^2) - 2x\sum_{\alpha=1}^{n}x_{i_\alpha} + \sum_{\alpha=1}^{n}x_{i_\alpha}^2 - 2y\sum_{\alpha=1}^{n}y_{i_\alpha} + \sum_{\alpha=1}^{n}y_{i_\alpha}^2. \quad (A.6)$$

So,

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\, f_{i_1}(x,y)\cdots f_{i_n}(x,y)$$

$$= \int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\, \exp\left[-Cn(x^2+y^2) + 2C\sum_{\alpha=1}^{n}(x_{i_\alpha}x + y_{i_\alpha}y) - C\sum_{\alpha=1}^{n}\left(x_{i_\alpha}^2 + y_{i_\alpha}^2\right)\right],$$

$$= \frac{\pi}{nC}\exp\left[\frac{C\left(\left(\sum_{\alpha=1}^{n}x_{i_\alpha}\right)^2 + \left(\sum_{\alpha=1}^{n}y_{i_\alpha}\right)^2\right)}{n} - C\sum_{\alpha=1}^{n}(x_{i_\alpha}^2 + y_{i_\alpha}^2)\right],$$

$$= \frac{\pi}{nC}\exp\left[-\frac{C}{n}\sum_{1=\alpha<\beta}^{n}\left((x_{i_\alpha}-x_{i_\beta})^2 + (y_{i_\alpha}-y_{i_\beta})^2\right)\right],$$

$$= \frac{\pi}{nC}\exp\left[-\frac{C}{n}\sum_{1=\alpha<\beta}^{n} d_{i_\alpha i_\beta}^2\right], \quad (A.7)$$

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

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A = \sum_{n=1}^{N_A}(-1)^{n+1}\sum_{1=i_1<\cdots<i_n}^{N_A}\frac{\pi}{nC}\exp\left[-\frac{C}{n}\sum_{1=\alpha<\beta}^{n}d_{i_\alpha i_\beta}^2\right], \quad (A.8)$$

$$= \frac{\pi}{C}\left[N_A - \frac{1}{2}\sum_{1=i_1<i_2}^{N_A}\exp\left[-\frac{C}{2}d_{i_1 i_2}^2\right] + \frac{1}{3}\sum_{1=i_1<i_2<i_3}^{N_A}\exp\left[-\frac{C}{3}\sum_{1=\alpha<\beta}^{3}d_{i_\alpha i_\beta}^2\right] + \cdots\right]. \quad (A.9)$$

Analogously we obtain the integral of the product of two membership functions of two species A and B,

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B = \sum_{n=1}^{N_A}\sum_{\tilde{n}=1}^{N_B}(-1)^{n+\tilde{n}}\sum_{1=i_1<\cdots<i_n}^{N_A}\sum_{1=\tilde{i}_1<\cdots<\tilde{i}_{\tilde{n}}}^{N_B}\frac{\pi}{(n+\tilde{n})C}$$

$$\exp\left[-\frac{C}{n+\tilde{n}}\left(\sum_{1=\alpha<\beta}^{n}d_{i_\alpha i_\beta}^2 + \sum_{1=\alpha<\beta}^{\tilde{n}}d_{\tilde{i}_\alpha \tilde{i}_\beta}^2 + \sum_{\alpha=1}^{n}\sum_{\tilde{\alpha}=1}^{\tilde{n}}d_{i_\alpha \tilde{i}_{\tilde{\alpha}}}^2\right)\right]. \quad (A.10)$$

10

B Toy models

B.1 $N_A = 1$, $N_B = 1$

We have in this case (here and in the following we indicate the indices - including the numbers - relative to the species B with a tilde),

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A = \int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_B = \frac{\pi}{C} \tag{B.1}$$

and

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B = \frac{\pi}{2C} e^{-\frac{C}{2}d_{1\tilde{1}}^2}. \tag{B.2}$$

being $d_{1\tilde{1}}$ the distance between the individuals of the species A and B.

B.2 $N_A = 2$, $N_B = 1$

We denote with a_1 and a_2 the two individual of the species A and with b the one of species B. Then, we have in this case

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A = \sum_{n=1}^{2}(-1)^{n+1} \sum_{1=i_1<\cdots<i_n}^{2} \frac{\pi}{nC} \exp\left[-\frac{C}{n}\sum_{1=\alpha<\beta}^{n} d_{i_\alpha i_\beta}^2\right], \tag{B.3}$$

$$= \frac{\pi}{C}\left[2 - \frac{1}{2}\sum_{1=i_1<i_2}^{2} \exp\left[-\frac{C}{2}d_{i_1 i_2}^2\right]\right], \tag{B.4}$$

$$= \frac{\pi}{C}\left[2 - \frac{1}{2}e^{-\frac{C}{2}d_{12}^2}\right], \tag{B.5}$$

being d_{12} the distance between a_1 and a_2, and

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A^2 = \sum_{n=1}^{2}\sum_{\tilde{n}=1}^{2}(-1)^{n+\tilde{n}} \sum_{1=i_1<\cdots<i_n}^{2} \sum_{1=\tilde{i}_1<\cdots<\tilde{i}_{\tilde{n}}}^{2} \frac{\pi}{(n+\tilde{n})C}$$

$$\exp\left[-\frac{C}{n+\tilde{n}}\left(\sum_{1=\alpha<\beta}^{n} d_{i_\alpha i_\beta}^2 + \sum_{1=\alpha<\beta}^{\tilde{n}} d_{\tilde{i}_\alpha \tilde{i}_\beta}^2 + \sum_{\alpha=1}^{n}\sum_{\tilde{\alpha}=1}^{\tilde{n}} d_{i_\alpha \tilde{i}_{\tilde{\alpha}}}^2\right)\right], \tag{B.6}$$

$$= \frac{\pi}{C}\left[1 + e^{-\frac{C}{2}d_{12}^2} - \frac{4}{3}e^{-\frac{2C}{3}d_{12}^2} + \frac{1}{4}e^{-Cd_{12}^2}\right]. \tag{B.7}$$

Also

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B = \sum_{n=1}^{2}\sum_{\tilde{n}=1}^{1}(-1)^{n+\tilde{n}}\sum_{1=i_1<\cdots<i_n}\sum_{1=\tilde{i}_1<\cdots<\tilde{i}_{\tilde{n}}}\frac{\pi}{(n+\tilde{n})C}$$

$$\exp\left[-\frac{C}{n+\tilde{n}}\left(\sum_{1=\alpha<\beta}^{n}d_{i_\alpha i_\beta}^2 + \sum_{1=\alpha<\beta}^{\tilde{n}}d_{\tilde{i}_\alpha \tilde{i}_\beta}^2 + \sum_{\alpha=1}^{n}\sum_{\tilde{\alpha}=1}^{\tilde{n}}d_{i_\alpha \tilde{i}_{\tilde{\alpha}}}^2\right)\right], \tag{B.8}$$

$$= \sum_{n=1}^{2}(-1)^{n+1}\sum_{1=i_1<\cdots<i_n}\frac{\pi}{(n+1)C}\exp\left[-\frac{C}{n+1}\left(\sum_{1=\alpha<\beta}^{n}d_{i_\alpha i_\beta}^2 + \sum_{\alpha=1}^{n}d_{i_\alpha \tilde{1}}^2\right)\right], \tag{B.9}$$

$$= \frac{\pi}{C}\left[\frac{1}{2}\sum_{1=i_1}^{2}\exp\left(-\frac{C}{2}d_{i_1\tilde{1}}^2\right) - \frac{1}{3}\exp\left(-\frac{C}{3}[d_{12}^2 + d_{1\tilde{1}}^2 + d_{2\tilde{1}}^2]\right)\right], \tag{B.10}$$

$$= \frac{\pi}{C}\left[\frac{1}{2}\exp\left(-\frac{C}{2}d_{1\tilde{1}}^2\right) + \frac{1}{2}\exp\left(-\frac{C}{2}d_{2\tilde{1}}^2\right) - \frac{1}{3}\exp\left(-\frac{C}{3}[d_{12}^2 + d_{1\tilde{1}}^2 + d_{2\tilde{1}}^2]\right)\right], \tag{B.11}$$

where $d_{i_\alpha \tilde{1}}$ is the distance between a_α and b.

- Suppose that $d_{1\tilde{1}} \ll 1$, namely that b is extremely close to a_1 (and consequently $d_{12} \simeq d_{2\tilde{1}}$). In that case,

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B = \frac{\pi}{C}\left[\frac{1}{2} + \frac{1}{2}\exp\left(-\frac{C}{2}d_{12}^2\right) - \frac{1}{3}\exp\left(-\frac{2C}{3}d_{12}^2\right)\right]. \tag{B.12}$$

The weight w_{AB} of the link $A \to B$ then results

$$w_{AB} = \frac{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B}{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A^2} = \frac{\frac{1}{2} + \frac{1}{2}e^{-\frac{C}{2}d_{12}^2} - \frac{1}{3}e^{-\frac{2C}{3}d_{12}^2}}{1 + e^{-\frac{C}{2}d_{12}^2} - \frac{4}{3}e^{-\frac{2C}{3}d_{12}^2} + \frac{1}{4}e^{-Cd_{12}^2}}, \tag{B.13}$$

and we see that the maximum value for w_{AB} is $\frac{2/3}{2/3+1/4} = \frac{8}{11}$ and the minimum value is $\frac{1}{2}$. The weight w_{BA} of the link $B \to A$ then results

$$w_{BA} = \frac{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B}{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_B^2} = \frac{\frac{1}{2} + \frac{1}{2}e^{-\frac{C}{2}d_{12}^2} - \frac{1}{3}e^{-\frac{2C}{3}d_{12}^2}}{\frac{1}{2}} = 1 + e^{-\frac{C}{2}d_{12}^2} - \frac{2}{3}e^{-\frac{2C}{3}d_{12}^2}, \tag{B.14}$$

with maximum value equal to 1 and minimum value equal to $\frac{4}{3}$.

- In the case that b is equidistant from a_1 and a_2, i.e. $d_{1\tilde{1}} = d_{2\tilde{1}}$, we have

$$\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B = \frac{\pi}{C}\left[\exp\left(-\frac{C}{2}d_{1\tilde{1}}^2\right) - \frac{1}{3}\exp\left(-\frac{C}{3}[d_{12}^2 + 2d_{1\tilde{1}}^2]\right)\right]. \tag{B.15}$$

The weights result to be given by

$$w_{AB} = \frac{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B}{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A^2} = \frac{e^{-\frac{C}{2}d_{1\bar{1}}^2} - \frac{1}{3}e^{-\frac{C}{3}[d_{12}^2 + 2d_{1\bar{1}}^2]}}{1 + e^{-\frac{C}{2}d_{12}^2} - \frac{4}{3}e^{-\frac{2C}{3}d_{12}^2} + \frac{1}{4}e^{-Cd_{12}^2}}, \tag{B.16}$$

with minimum and maximum values respectively equal to 0 and $\frac{2/3}{11/12} = \frac{8}{11}$; and

$$w_{BA} = \frac{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_A\mu_B}{\int_{\mathbb{R}^2} \mathrm{d}x\,\mathrm{d}y\,\mu_B^2} = \frac{1}{2}e^{-\frac{C}{2}d_{1\bar{1}}^2} - \frac{1}{6}e^{-\frac{C}{3}[d_{12}^2 + 2d_{1\bar{1}}^2]}, \tag{B.17}$$

with minimum and maximum values respectively equal to 0 and $\frac{1}{3}$.