Description of the model

We formulate our model as a directed graph consisting of N nodes. The connections between nodes form a network. This network can be represented with an N x N connectivity matrix, $A = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 : \text{if } i \text{ connects to } j \\ 0 : \text{otherwise} \end{cases}$$

Each network configuration or *topology* (up to renumbering of the nodes) will represent a different organism and or energy/signal network.

We will assume that one node in the network receives a *signal* (energy) from the environment, and we will call this node the *pump*. From the pump, the signal (energy) is transmitted along edges to all other nodes in the network, with each directed edge representing signal transmission . Heretofore, the pump will be labeled as node 1 in the connectivity matrix, and in diagrams as P.

All nodes can receive a signal along each incoming edge and transmits a signal along each outgoing edge. Each node has a constant *signal difference*, c, that is the difference between the total signal received and the total transmitted signal: $c = E_{in} - E_{out}$ (note if $c > 0$ that corresponds to having a certain energy cost per cell) [NOTE: non-constant c is possible in extensions].

The network of nodes that we will consider are directed graph that contains a directed path from the pump to any other node in the network, we will refer to this condition as *pump-connectedness*].
After the *signal difference* is applied, each node i divides its remaining signal $(E_{out,i})$(energy) into equal parts based on the total number of outward connections, b_i. The amount received by each of the b_i successors is equal to a proportionality factor, p, of the outgoing energy:

$$E_{in,j} = \frac{E_{out,i} \cdot p}{b_i} \tag{0.1}$$

There are three cases to consider, $0 < p < 1$: dissipative transport, the signal is reduced during transport; $p > 1$: amplified transport, the signal increases during transport; and $p = 1$: neutral transport, the signal is unchanged during transport.

The signal (energy) at each node will then be given by the sum of the signal received from inward connections. We will denote the signal transmitted from node i to node j as E_{ij}. Therefore, given an initial signal (or environmental energy), E_0, we can calculate the energy at each successive node i using the formula:

$$E_i = \sum_{j=1}^{N} E_{ji} = \sum_{j=1}^{N} a_{ji} \frac{(E_j - c)p}{b_j}, \text{ For } i = 2, \cdots, N \tag{0.2}$$

$$E_1 = E_0 + \sum_{j=1}^{N} E_{j1} = \sum_{j=1}^{N} a_{j1} \frac{(E_j - c)p}{b_j} \tag{0.3}$$

where E_i is the total incoming signal to node i (Note: $E_i = E_{in,i}$ for node i) for $i = 1, 2, ..., N$ and E_0 is the environmental (initial) signal given to the system from an outside source [NOTE: May want to change notation to be more graph theoretical (i.e. $j \in V(G)$)]. An example of the use of 0.2 and 0.3 for a simple system is shown in figure 1[create reference].
Note, that if instead we know the signal at one specific node i, using the previous linear system we can identify all the other signals, including the initial signal E_0 required for node i to have that prescribed energy E_i. We will use this to identify important initial signal levels for the network.
Simple examples (figure 1)

The critical energy of the k^{th} node E_k^* is the lowest environmental energy at which the node can survive (i.e. $E_k \geq c$) (note, it is possible to have a critical value for death other than c)
To find this energy we take the previous linear system of N equations and add the condition $E_k = c$ so that we have $N + 1$ equations in $N + 1$ unknowns. Generally this system has a unique solution since the equations are linearly independent (needs prove?). Therefore, we can solve for the E_i with $i = 0, 1, ..., N$ and we have $E_k^* = E_0$.

Once we have a list of critical energies for all nodes, the first node to die will be the one with the largest critical energy, then the one with the second largest critical energy, and so on. Therefore, forming a list of critical energies in descending order, will correspond to have a list of death events, where every time the energy falls below an entry a certain number of nodes will die, corresponding to the position in the list of the entry. The r^{th} entry in this list will be called the r^{th} critical energy and denoted $E^{*,r}$.
Note however then when calculating critical energies we should also take into consideration that nodes that receive less than c amount of energies cannot transmit to the next node, effectively changing the system to solve (this is a simple elimination of terms in the previous equation and can be easily implemented).

3. Special Networks

Therefore, using our explicit formula we can solve the system and eventually write down the value of E_0 as a function of the other nodes energies, and by substitution obtain an explicit formula dependent on c and p. However, for most topologies this process is very specific and knowing the explicit formula for one does not generally allows to express the same for an other topology. Nevertheless, we can group topologies into ones that have common characteristic for which we can explicitly write the formula for critical energy. (Need to reformulate in terms of extendability: A sufficient condition to write the critical energies is for the topology to be acyclic). Some examples of such topologies are seen in figure 2 [reference figure. To include: Description of special topologies in caption, Topologies themselves, connectivity matrix, critical energies] [Followed by plots: .....]. For most of the paper from now on we will consider $c > 0$ and $0 < p < 1$ (both fixed), but in the following sections

we will also briefly discuss other cases. This setting is representative of a biological network in which nodes need to consume energy to stay alive ($c > 0$) and moving energy between nodes leads to a loss of energy with a proportion p. We will now illustrate in some detail some of the topologies previously mentioned. Starting with the simplest *star* and *line*, and then moving to slightly more complicated groups that we will call *forks* and *flares*. For this topologies finding the first node to die is relatively simple and therefore reconstructing an explicit formula for its critical energy its a simple reverse calculation. We will consider for this only the 1st critical energy, since upon one death this topology still become topologies of one of the described type.

a. Line

A line topology is one in which each node (except the last one, first to die) connects to one and only one other node and each node (except the first one, pump) is connected to one and only one other node (see fig.2). This topology maximizes the number of connections to reach the last node (first death), and is therefore the one (when the size of the network is fixed) that requires the highest energy to keep every node alive (note if $p > 1$ the opposite is true).

b. Star

In a star topology the pump connects to all other nodes and other nodes have no outgoing connections. In this case all non-pump nodes will die at the same time, but the number of connections from pump to them is minimized to 1 making this the topology with the lowest energy to keep everyone alive.

c. Importance of these motifs

As we can see this two topologies despite beeing the simplest correspond to the two extreme points in terms of critical energies, being the best and the worst. We then want to preserve some of their characteristic while at the same type producing hybrid topologies that are partially line-like and partially star-like.

d. Simple generalizations

To do so we will explore two possible approaches, in which we have topologies with only one node with multiple connections, but only at one of the two extremity of the energy transfer (an even further generalization with the bifurcation at any point on the longest line-path is presented in the supplementary material(?)).

i. Fork

A fork is a topology that extends as a line up to the l-th node and then this node has s connection one to each remaining node (see figure 2) (note that $l + s = n$ total number of nodes, and $l, s \geq 1$).

ii. Flare (Stellar flare (re-name))
A flare is a topology in which the pump has exactly s outgoing connections and one of the branches formed this way has exactly length l (not counting the pump) (note that $l + s = n$ total number of nodes, and $l, s \geq 1$).

Note that for both flares and forks $l + s = n$ and these parameters can be used to indicate how close they are to be a line or a star, with $s = n - 1$ and $l = 1$ being a star, and $l = n - 1$ and $s = 1$ being a line. (more explanation in figure caption??)

e. Provide explicit formula of first critical energy
In figure 2 we are summarizing these topologies and their explicit formulas for critical energy, while in figure 3 we are analyzing in details what happens to the critical energy for the flares when we fix p, c, n and vary s. As we can see in this case depending on the environmental energy, both being close to a line or a star can be beneficial, but not being in an intermediate condition (more details in figure 3, parabola, net graphs, etc...)(should we include colormaps with only these special topologies for a certian size?).

f. Node to die is independent of c and p
In all the previous topologies it is easy to find the first node to die and calculate its energy based on E_0, then imposing that this energy is exactly equal to c we can find the corresponding critical energy. For the star and fork all the nodes after the bifurcation will have the same critical energy and will die first since they are the last to receive energy. while for line and flares only the final node with no outward connection will die at the first death event. Note that while stars reduce the required energy to keep everyone alive, they lose $n - 1$ nodes all at once, while lines lose one node at a time. Flares and forks have instead an intermediate behavior, with the first losing one node at a time for the first $l - 1$ nodes and then s all at once, and the fork loses first s nodes all at once and then $l-1$ nodes one at a time.

i. only one path back to the pump
This topologies are also particularly easy because for each node they have only one directed path connecting them to the pump, making the process of finding an explicit formula easier. However, the same process we used can be accomplished even with more paths (add example in the supp material?).
One other important characteristic is the absence of loops that would make our explicit formula an infinite series. If we have loops it is still possible to write an explicit formula, but we need to be careful and check that our series converges (add figure to explain non converging case?).

Simulations (should we call them simulations?)

Now that we have defined the critical energies as metrics to evaluate the success of topologies, and we have shown how simple topologies behave with respect to the first ordinal critical energy, that we will call just critical energy from now on, we want to ask ourselves questions about this "metric" for some simple cases and

then we will try to extract some information from more complicated one through simulations and analysis of these through statistical measures.

First we want to show some interesting properties that can be seen when comparing some relatively simple topologies, then we will give a general rule to identify the "best" topology, and finally we will show results for a big number of topologies and some interesting statistical measure about their critical energies.

4 topologies 5 nodes example / Coexistence and Conditional higher fitness

In figure 4 (colormap with 4 topologies) we are presenting four topologies all of size $n = 5$ and we are plotting how many of them survive (do not lose nodes) for different values of E_0 and p. In this simple cases we can still write down explicit formulas for the critical energy, and doing so we can verify that two of them have the same critical energy (bottom curve), and the other two have different critical energies with an intersection for $p = 1/3$ (actual calculations shown in supp. material?).
Therefore, this simple example shows us that we can have networks with different topologies being equivalent in terms of critical energy, and in the second case different networks having critical energies for which one is lower than the other for certain values and higher for other values.
We can think of the first phenomenon as the coexistence of two different species with the same fitness, and the second one as two species whose fitness relative to each other changes dependent on p or E_0.
Moreover after observing this example is easy to imagine that more complicated situation could arise when considering more complex network, as coexistence of more than two network, two networks with multiple intersections, intersection in one point of more than two critical energies curves, etc... However, trying to classify all the possible situation is not the main purpose of our paper. Instead we will now give a condition to be the "best" topology and relate this to the previous example.

"Best" topology

We have already established that when considering $0 \leq p \leq 1$ and $c > 0$ the star topology is the one with the lowest $1st$ critical energy for any specific size n. Now we would like to state a general rule for the other ordinal critical energies.
Lets take for instance the $2nd$ critical energy for a network. This means that we want to find the network of a fixed size n whose second critical energy is the lowest possible. This critical energy is by definition the minimum energy before which we get a second node to die, lets assume for now that 1st and 2nd critical energies are different. Therefore, we should consider a network of $n-1$ nodes with lowest $1st$ critical energy and add one node in such a way that the added node will die before any of the $n-1$ nodes. However, we already know that for networks of a certain size the topology with the lowest 1st critical energy is the star. This then implies that the network we seek is a star with $n-1$ nodes with an additional node connected to at least one of the $n-2$ ramification from the center, and by construction the 2nd critical energy of this network is the same as the 1st critical energy of the star with $n-1$ nodes.

Lets now consider the case in which 1st and 2nd critical energy are the same. If the best 2nd critical energy topology had this property then the 2nd critical energy would be equal to the 1st critical energy and then will be less or equal to the first critical energy for the star with n nodes, but this is obviously bigger than the 1st critical energy of the star with $n-1$ nodes, hence topologies with this property have higher second critical energy than the one described above.
All this process can be easily extended to any ordinal critical energy. Lets say that we want to minimize the r-th critical energy, then we will consider a star with $n-(r-1) = n-r+1$ nodes and will then add $r-1$ nodes to it in such a way that this will die before any of the other $n-r+1$.
It is easy to see that if we are minimizing only one critical energy, other than the first, there are many topologies that will work, each with a common part (star with $n-r+1$ nodes), and the higher the order of the critical energy the more topologies we get. One alternative would be to specify critical energies to minimize one after an other from the higher order to the lower one. We are providing some example of minimized topologies in figure 5 (see supp. material for more details).

Our next step will be to compare a big number of topologies together through their critical energies. However, is not hard to see that even with our condition of pump-connectedness and even considering the isomorphic topologies (the same up to renumbering of nodes, and/or elimination of unused connections) when we consider a decent topology size $n > 5$ the number of possible topologies grows in a factorial fashion. Therefore, to obtain a useful sample of this big set of topologies and being able to compare their critical energies (we are doing 1st, but others could be done) we are either using "random" topologies or "semi-optimal" topologies (for more details see supplementary material). Moreover, we need to remember that if $p > 1$ and/or $c < 0$ our solutions for the critical energy can result in divergent series. This happens when we have cycles in the topologies (see fig. 6 for an example).
In our study we will consider samples of topologies all of the same size (example $n = 10$), topologies not larger than a certain size (ex $n \leq 10$), and topologies of similar size (example $n = 9, 10, 11$).
When comparing topologies of different size we will consider the environmental energy per cell as variable instead of the environmental energy, so that each topology will receive energy proportional to its size (E_0/n is constant for each point y- cross section in our colormaps).
Our colormaps (see figure 7) show in color-scale how many topologies in the sample will survive (lose no nodes, 1st crit. energ.) for certain E_0/n and p values in the plotted range (log10 scale for energy). Careful observation will show that we still have intersection and/or overlapping as seen in our simple example, but the number of topologies makes it unreasonable to find all of them.
What we will do is to underline features of these plots reading them in different ways and extracting some statistical values from our sample.

i. Evolvability window

Our first quantity of interest is the range of energies for which some but not all topologies can survive, in other words the energy values for which we have at least

one topology surviving and at least one dying, we will indicate the number of surviving topology with N and the total number in our sample with N_{max} and we will call this range of energies *evolvability window* or *directed-evolution window*. This name comes from the fact that for those energies there is real competition, since we have that at least one survives when the energy is distributed equally, but not everyone. Our first observation is that the size of this window changes with p and in particular is always decreasing for $0 \leq p \leq 1$ (fig.8). However, we have two different behaviors if we consider topologies of many different sizes ($n \leq 10$) or only topologies of approximately the same size. In the first case the window size always decreases, till it reaches a value approximately equal to zero (somewhere after $p > 1$ see fig. 9), while in the second case it reaches a minimum value (for some $p > 1$ but close to $p = 1$ fig 10, need to retest) and then becomes slightly increasing. This behavior is even better represented through p-cross-sections from our colormaps (see supp. material).

ii. Survivability mean and standard deviation, with respect to p

We then wanted to observe how the number of surviving species changes with respect to our transport efficiency parameter p, with reasonable environmental conditions. For this purpose we are considering an environment where the energy has an entropy maximizing probability distribution (see supp. material for details) with a certain mean (in fig. 11 and 12 we are considering 5 different mean energies) and for each p-cross-section we are calculating mean and standard deviation for the number of surviving topologies in that section based on the probability of the corresponding energies (see fig. 11 and 12).
As expected the means are increasing functions of p and tend to N_{max}. However, the slope of this curves drastically changes from 0 at $p = 0$ to a certain maximum value and then decrease again till 0 when the curve reaches N_{max} (need to use low enough mean energy to be able to see this effect). Therefore, for each curve (different mean energy), we have a value of p for which the average change in survivability is maximum.
On the other end the standard deviations show that for low p as well as high p the uncertainty of survivability is low while there is a p value for which the uncertainty is at its highest.
We can see that ideally a topology would want to be close to the maximum slope for the slope of the mean survival, but away from the maximum for the standard deviation, that indicates uncertainty.
We also derived analogus graphs for the E-cross-sections, with similar results.
These result will be further explain in our conclusions.

Change in peaks for slope of mean and std survaivability for different sizes
To show emergence of maximum size?(work in progress)

Comparison with PDE model on 2D or 3D domain
(work in progress)

Conclusions

Through our simple network model we provided a simplified way of looking at energy exchange for biological organisms (considered as collections of cells) or exchange of signals in a connectivity network, while at the same time we where able to underline a number of features observed in real organism seeking an energy source to keep their cells alive.
With simple example we showed how branching (star) is preferential to line structures, but also that in certain conditions more line-like structure are acceptable, and that intermediate structure can be worst than any of the two extremes.
We have also seen how the same "solution" (critical energy) can be obtained with two different strategies (topologies), similarly to two organisms independently evolving in the same environment to a common fitness level producing a coexistence situation.
We were also able to see how an improvement in efficiency actually corresponds to a decrease in competition and therefore leads to evolutionary plateaus.
Moreover, through statistical analysis of our simulations we were able to determine a preferential window for our efficiency parameter p (work in progress) based on standard deviation and means extracted from our colormaps, where the minimum value correspond to the highest uncertainty (std peak) of survival and the highest to the last point at which increasing p is considered advantageous compared to its cost.
We also showed how maximum sizes naturally arise under certain environmental conditions, and we explained this effect through probability measures, since having optimal or even good topologies its statistically harder for bigger networks.
Finally we have shown how our model can also be considered an approximation of a reaction-advection-diffusion PDE model in certain 2D or 3D domains with certain parameters corresponding to our p, c (work in progress).

Future directions (???)

FIGURE 1. Example of use of implicit formulas

FIGURE 2. special topologies table: line, star, fork, flare

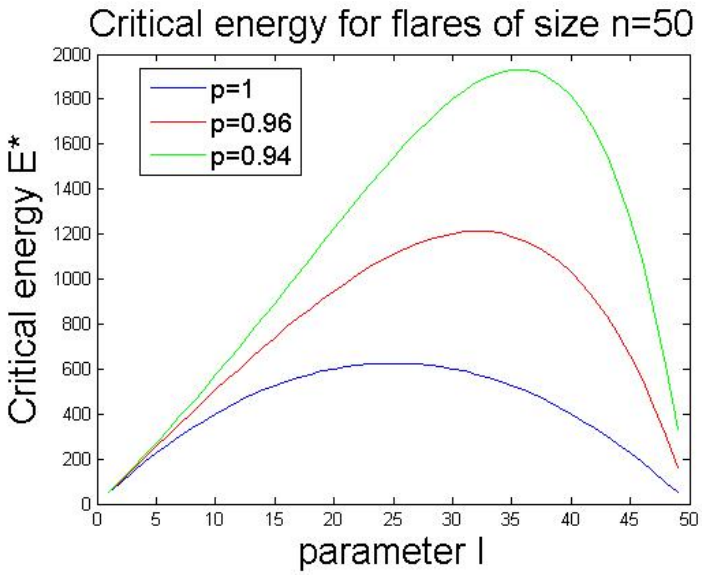

FIGURE 3. Plots of first critical energy for flares of size $n = 50$ with respect to prameter l, $c = 1, p = 1, 0.96, 0.94$

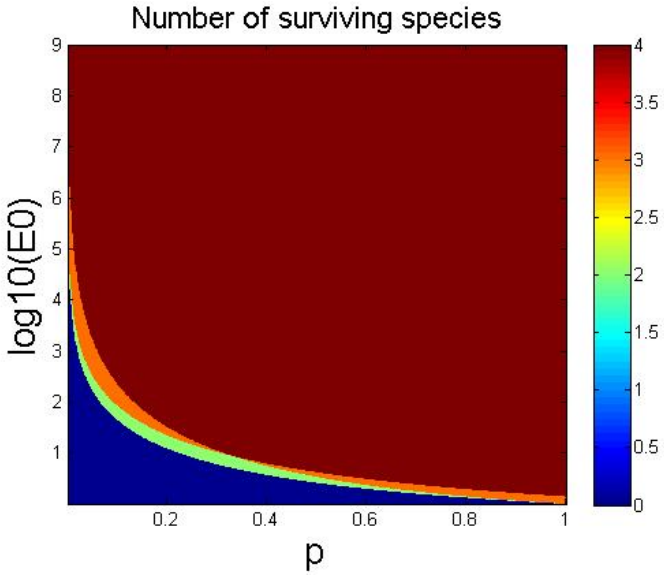

FIGURE 4. Number of surviving species for 4 topologies of size 5

FIGURE 5. minimized topologies

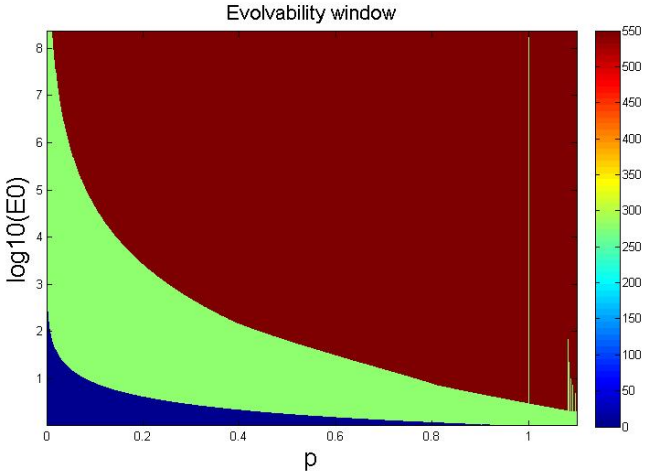

FIGURE 6. colormap of topologies with cycle and p¿1

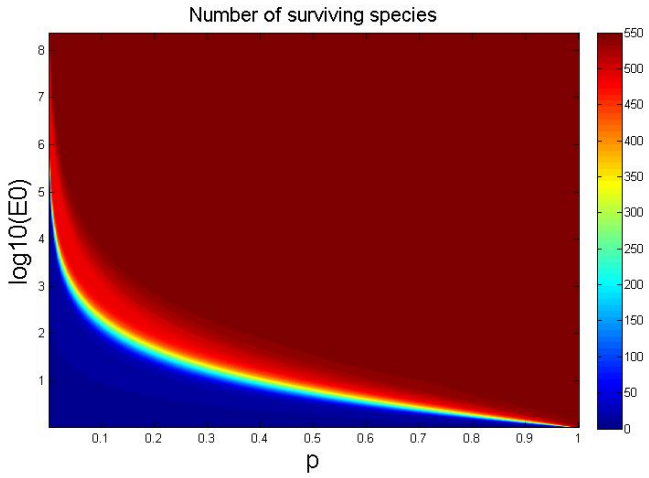

FIGURE 7. Surviving topologies. Sizes $n = 5, .., 15$,with 50 each, random topologies

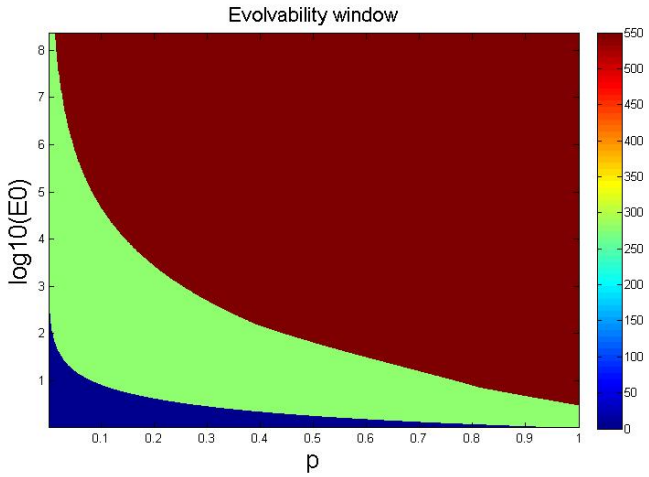

FIGURE 8. Window of evolvability

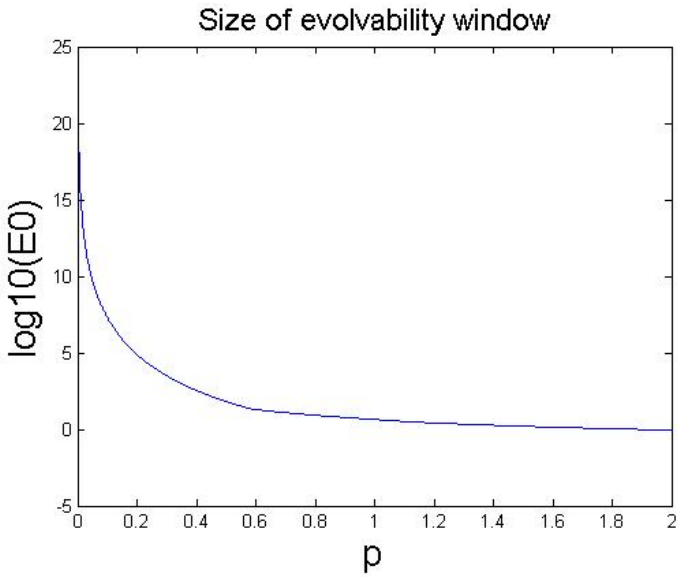

FIGURE 9. size of window of evolvability for $0 \leq p \leq 2$ no cycles $n = 2, ..10$, with total 3741 topologies (builder)

FIGURE 10. size of window of evolvability for $0 \leq p \leq 2$ size $n = 10$. (to redo)

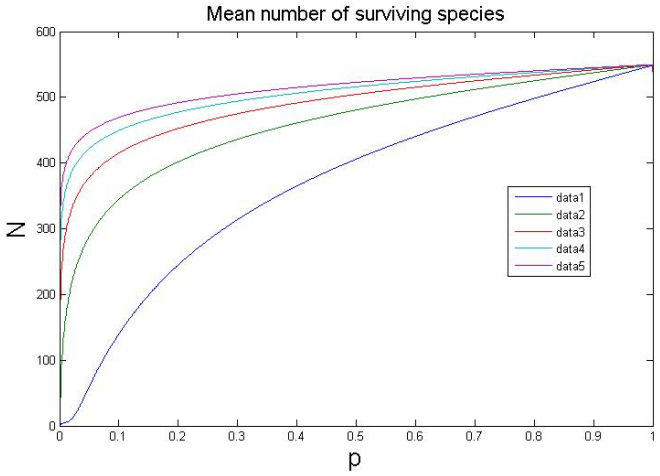

FIGURE 11. mean number of surviving species for different mean energies (lowest to highest), as a function of p. Sizes $n = 5, .., 15$,with 50 each, random topologies

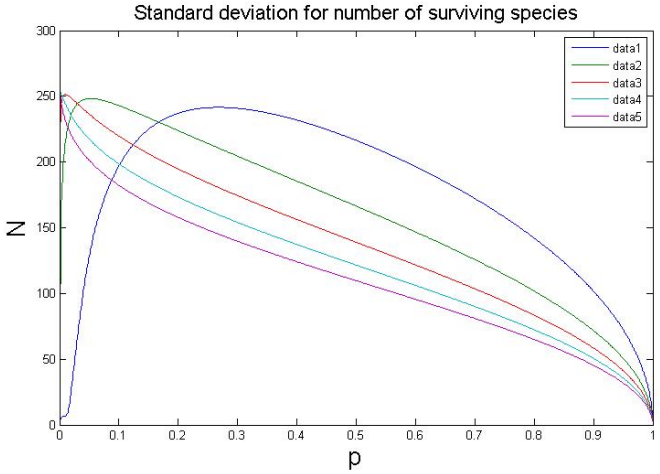

FIGURE 12. standard deviation of number of surviving species for different mean energies (lowest to highest), as a function of p. Sizes $n = 5, .., 15$,with 50 each, random topologies