On the edge of forever - some questions about the complete redirection network

Marcos Costa Santos Carreira*

06-Oct-2022

Abstract

Using the complete redirection model from [KR2017] and a more detailed classification of nodes, we look at the general redirection networks and ask the first question: is there a redirection probability lower than 1 that leads to the same kind of sublinear growth for the nucleus?. We look at the distribution of degrees for the nucleus as restricted $(H_j, C_k > 1)$ partitions of $2\,(N-1) - L$ with $N - L$ elements, highlighting the difference between hubs and cores. We then look at the 2 feedback components of the redirection network: (i) the direct feedback, due to the unbalanced dynamics of the redirection and (ii) the geometry of the redirection, where hubs with more leaves have a higher chance of sprouting a new leaf compared with hubs with less leaves. After simulating a redirection network where the redirection is global (according to the distribution of elements) minimizing the effect of the geometry, we ask the second question: is the geometry component necessary for the nucleus fraction to vanish? We think the answers are No and Yes, respectively, because of the same reason: the "drift" towards an expected value of 0 on the complete redirection network is just enough to do it; any counterflow will disrupt it.

*École Polytechnique, CMAP - PhD under the Quantitative Regulation chair

1 Introduction

Steven Strogatz posted the link to [KR2017] in 20-Sep-2022 when he was visiting the Santa Fe Institute. The problem of finding why the number of nucleus nodes $N_{nucl} = \sum_{k\geq 2} N_k$ grows as N^{μ} with $\mu \approx 0.566$ and the predominance of star-like networks is really interesting, and this is what we are going to explore. In a previous short paper ([C2022]) we explored just the complete redirection network. We will incorporate the previous paper in the present paper.

2 Model description

2.1 Rules

2.1.1 Complete Redirection

The complete redirection model for undirected networks can be defined by the following rules (given an initial connected graph):

1. A new node chooses a provisional target node uniformly at random
2. A neighbor (a node connected to the provisional target) is chosen uniformly at random among all the neighbors as the attachment point of the new node

Figure 1 is an example of such an iteration:

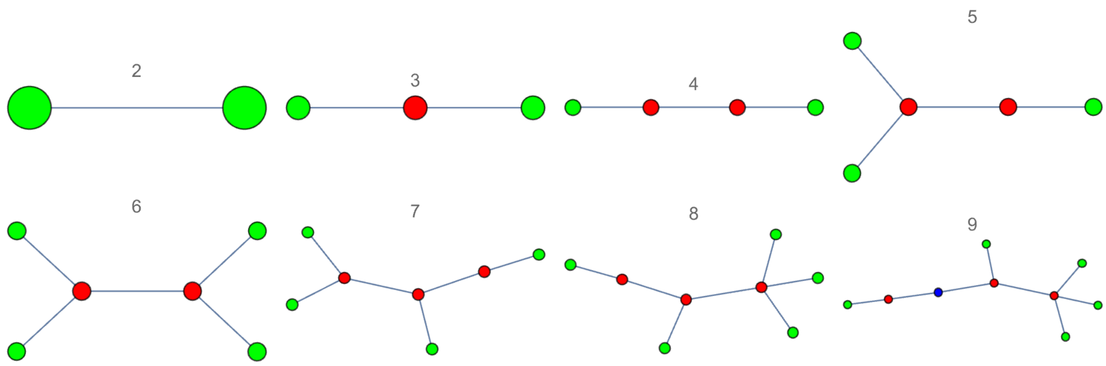

Figure 1: A short iteration; leaves=green, hubs=red, cores=blue

2.1.2 "Fuzzy" Complete Redirection

The "fuzzy" complete redirection model is similar to the complete redirection model, but without the geometry of the connections:

1. A new node chooses a provisional target node uniformly at random
2. Neighbors' choice:
   (a) If the node is a leaf, the neighbors are considered to be all the hubs (see the classification below)
   (b) If the node is a hub, all the other nodes are eligible as neighbors

(c) If the node is a core, the neighbors are considered to be all the hubs and the other cores (no leaves)

3. The attachment point of the new node is chosen uniformly at random among all the neighbors

2.1.3 No Redirection

This is simple:

1. A new node chooses a target node uniformly at random and attaches to it
2. A neighbor (a node connected to the provisional target) is chosen uniformly at random among all the neighbors as the attachment point of the new node

2.1.4 Partial redirection

The partial redirection depends on the probability pr of redirection;

1. A new node chooses a provisional target node uniformly at random
2. A random real number between 0 and 1 is chosen
   (a) If it is lower or equal than pr, the complete redirection follows from step 2
   (b) If it is greater than pr, the node chosen is the attachment point (as per the no redirection rule)

2.2 Classifications

We classify the nodes as:

1. Leaves: Exactly 1 edge, connects to a hub; every new node starts as a leaf
2. Hubs: More than 1 edge and connects necessarily with at least 1 leaf; can be connected with other hubs and with cores
3. Cores: More than 1 edge and it doesn't connect with leaves, only with hubs and other cores

So we have our first equation relating the number of nodes and their classification:

$$N = L + H + C \tag{1}$$

And a second equation relates the number of edges E and the connections among the different components:

$$E = N - 1 = L + HH + HC + CC \tag{2}$$

Where Leaf to Hub (LH) is, by definition, equal to L.

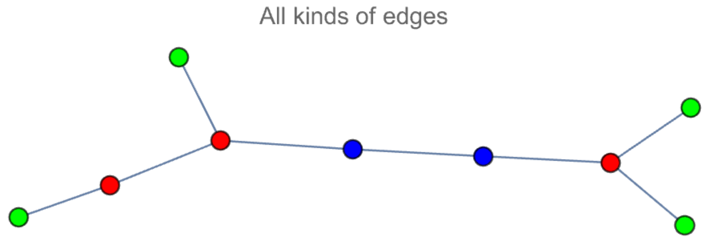

Figure 2: A network example; leaves=green, hubs=red, cores=blue

The last network of Figure 2 , with N=9 and E=8, has (L, H, C) = (4, 3, 2) and (L, HH, HC, CC) = (4, 1, 2, 1).

2.3 Iterations and Nucleus frequency

We can simulate the growth of the network (even with different initial networks, with similar results) and monitor the relative frequency of nucleus elements:

$$N_{nucl} = \sum_{k \geq 2} N_k = H + C \tag{3}$$

And plot it over time, in logscale (Figure 3):

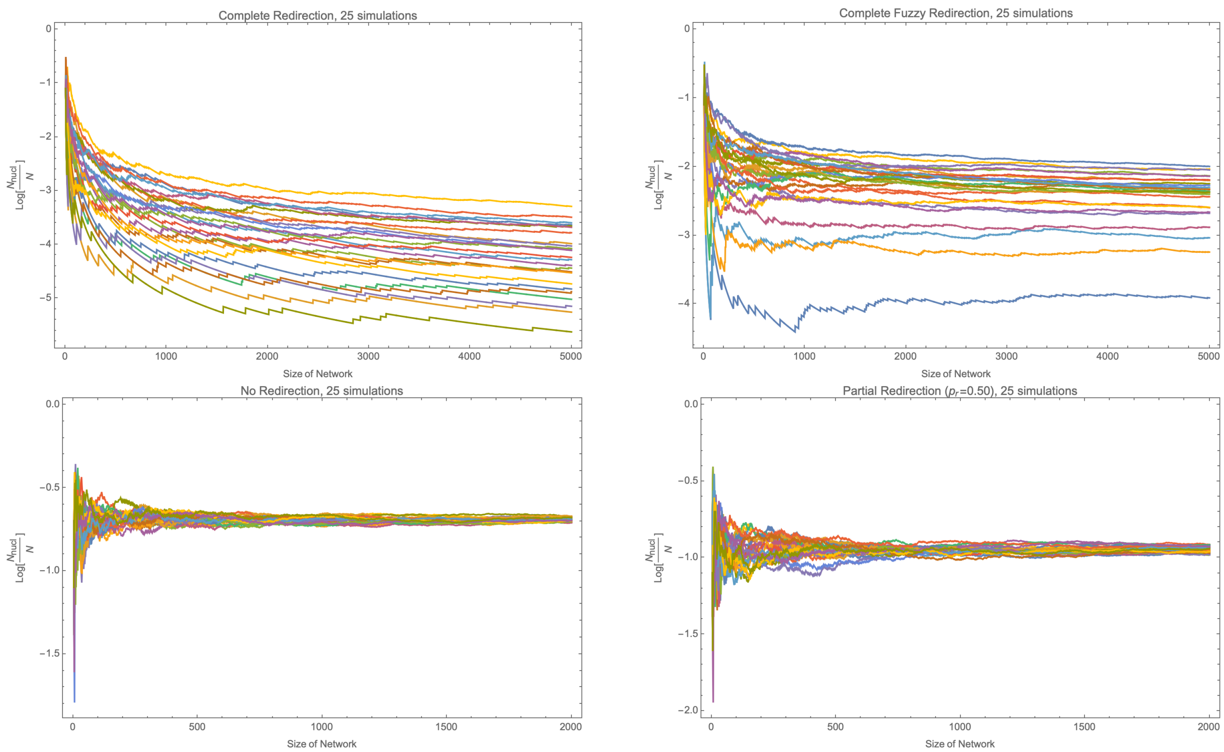

Figure 3: Nucleus frequency over 25 simulations

3 Automating the iterations

3.1 Graphs

We now must span all possible node choices and, for each choice, all neighbors. We must then tally the frequencies of isomorphic networks.

Fortunately, Mathematica has the CanonicalGraph command, which enables us to collect Graphs with the same structure but different orderings for the labels.

The following Figures (4, 5, 6 and 7) show the probabilities for each network in the complete redirection model up to N=9:

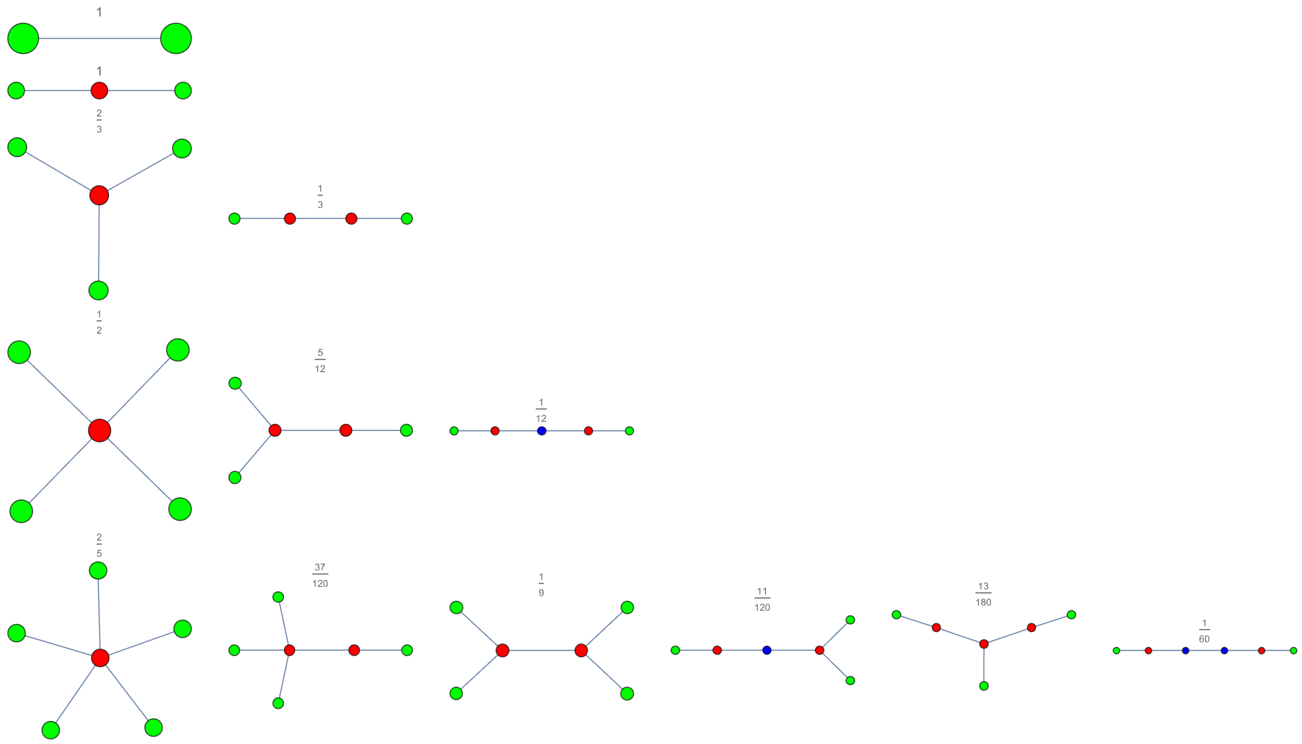

Figure 4: Enumeration of all networks configurations up to N=6

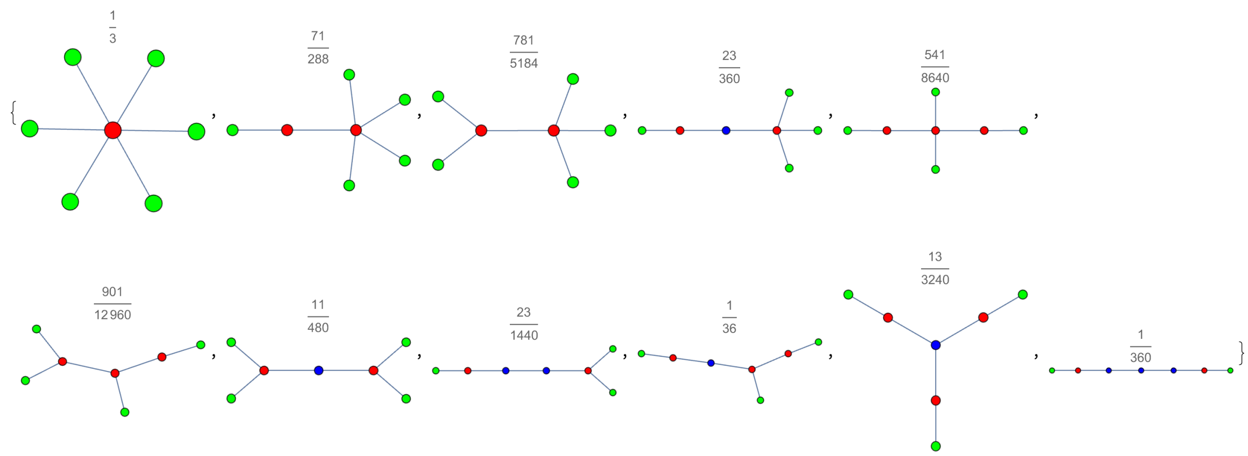

Figure 5: Enumeration of all networks configurations for N=7

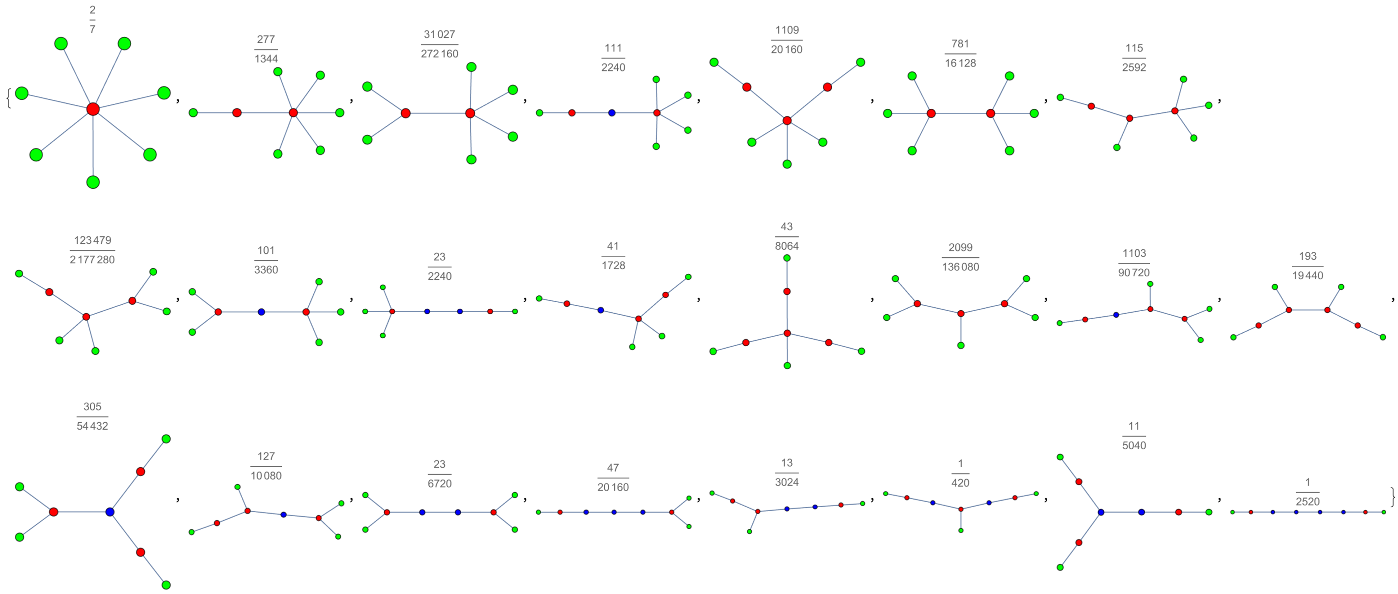

Figure 6: Enumeration of all networks configurations for N=8

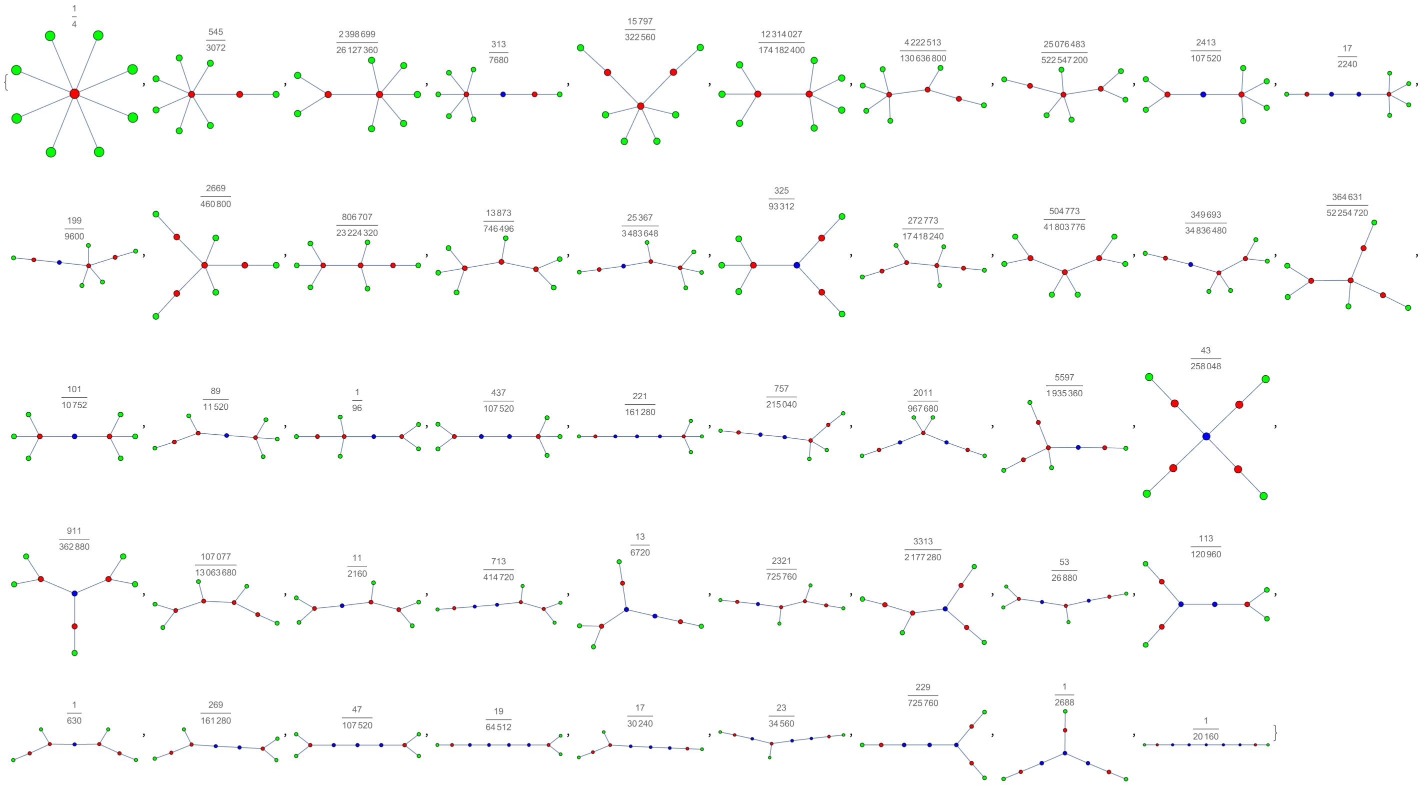

Figure 7: Enumeration of all networks configurations for N=9

For the other models we have the same structures, but with different probabilities.

3.2 PDFs

We grouped the networks of the complete redirection model by the number of leaves and looked at the PDFs for N up to 16 (Figure 8). We can see how the extremes of each PDF slide down over the boundaries (Snake and Star), and of course the left side (few leaves) falls much faster (factorial) than the right side (more leaves) does (linear). This resembles a bedroll unfurling with a fast push down on the left side. at every step the PDFs is lower and the N-k leaves spiral up and then down, always to the right.

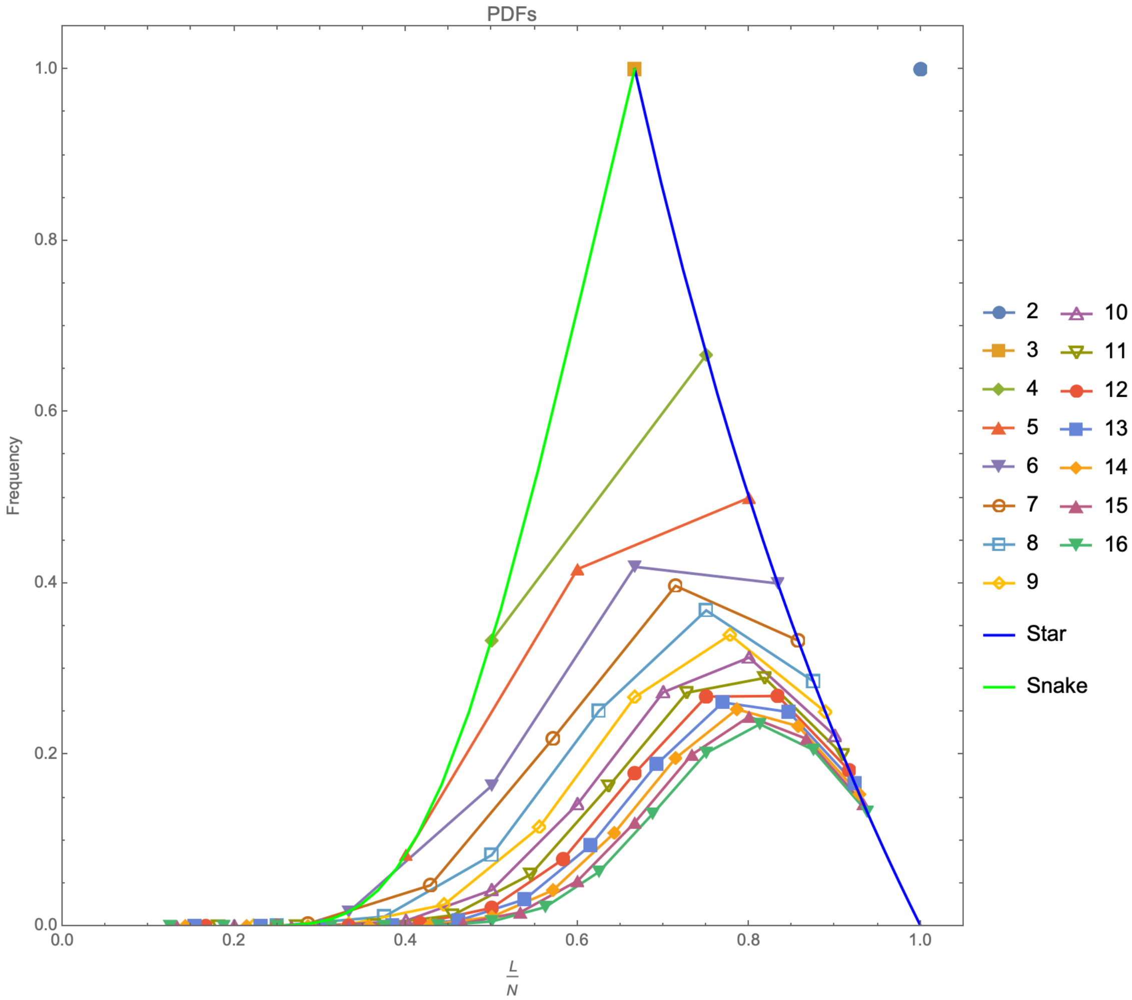

Figure 8: PDFs for $\frac{L}{N}$ up to N=16

Showing only the PDF for N=16 (Figure 9):

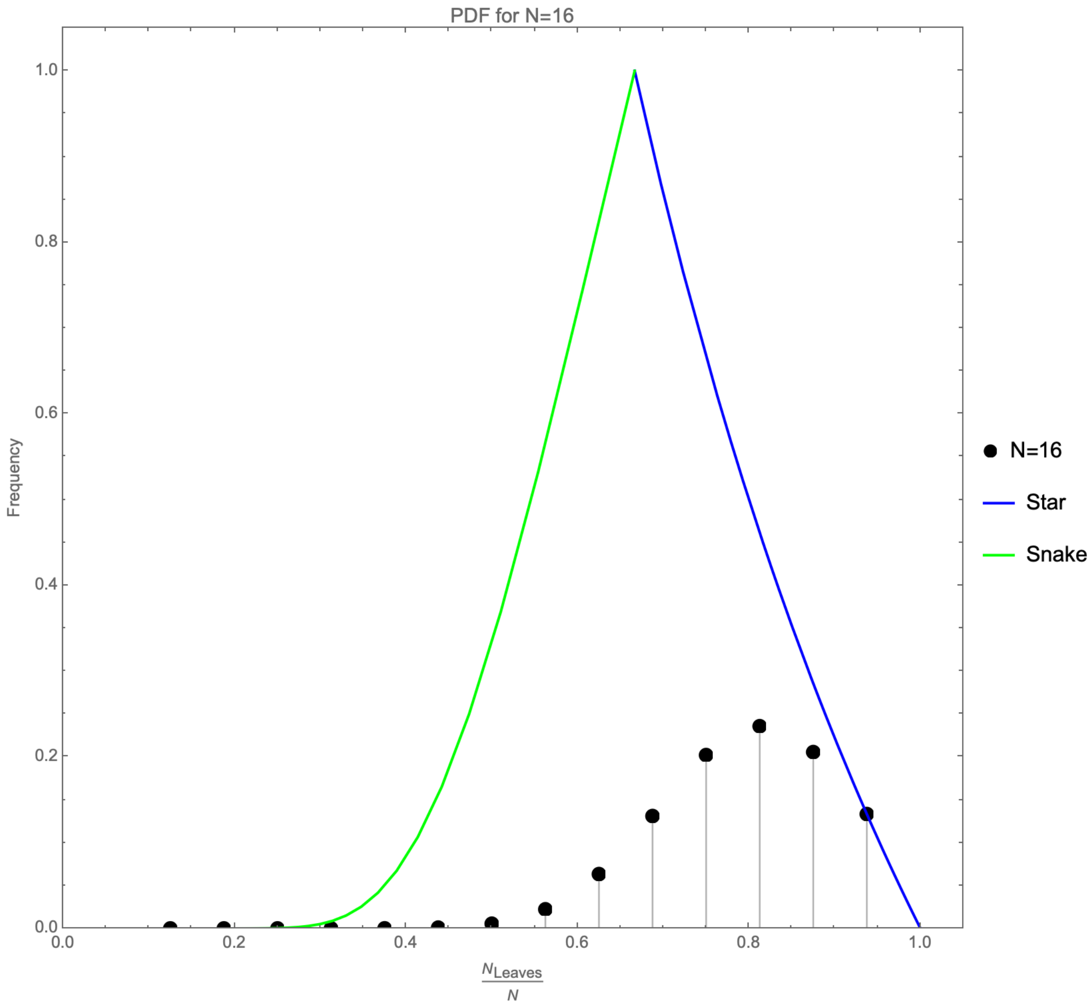

Figure 9: PDF for $\frac{L}{N}$ with N=16

Even if we change the initial conditions to a closed graph we get a similar resulted (truncated on the right but the mass always shifts to the right).

Each model will have its own PDF dynamics (more on this later).

4 Exact formulas for the complete redirection model

4.1 Perfect star

Starting from N=3, where (L, H, C) = (2, 1, 0), we must always choose any of the leaves $\left(\frac{N-1}{N}\right)$ so the hub will spring a new leaf:

$$P_{Star}\left[N\right] = \frac{2}{N-1} \tag{4}$$

4.2 Snake (line)

Starting from N=3, where (L, H, C) = (2, 1, 0), we must always choose any of the hubs $\left(\frac{2}{N}\right)$ and each hub will have a $\left(\frac{1}{2}\right)$ probability of choosing a leaf:

$$P_{Snake}\left[N\right] = \frac{2}{(N-1)!} \tag{5}$$

4.3 Single-defect star

The smallest single-defect star could be defined as the N=4 Snake (a leaf became a hub). But we will focus on the next single-defect star (Figure 10), with N=5 and (L, H, C) = (3, 2, 0):

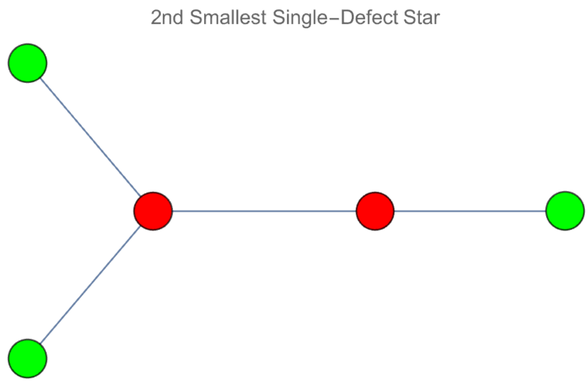

Figure 10: Single-defect star

If we already have a SD star with N nodes ($P_{SD}\left[N\right]$), we must either choose any of the leaves that connect to the hub with more leaves $\left(\frac{N-3}{N}\right)$ or the other hub $\left(\frac{1}{N}\right)$ followed by choosing the first hub $\left(\frac{1}{2}\right)$ so that the first hub can spring a new leaf.

We can also get a SD star from a perfect star $\left(\frac{2}{N-1}\right)$ and choosing the hub $\left(\frac{1}{N}\right)$. Then we can solve the recursion below for $N \geq 5$:

$$P_{SD}\left[N+1\right] = \left(\frac{2}{N\left(N-1\right)}\right) + P_{SD}\left[N\right]\left(\frac{N-3}{N} + \frac{1}{2N}\right) \tag{6}$$

Assuming:

$$P_{SD}\left[5\right] = \frac{5}{12} \tag{7}$$

Which can be solved by Mathematica and expressed as:

$$P_{SD}\left[N\right] = \frac{4}{3}\left(\frac{1}{N-1} + \frac{2}{\sqrt{\pi}}\frac{\Gamma\left[N-\frac{5}{2}\right]}{\Gamma\left[N\right]}\right) \tag{8}$$

Or getting rid of $\sqrt{\pi}$:

$$P_{SD}\left[N\right] = \frac{4}{3\left(N-1\right)}\left(1 + \frac{2^{8-2N}}{(N-2)}\binom{2N-7}{N-4}\right) \tag{9}$$

And we can confirm that:

$$\lim_{n\to\infty}\left[\frac{P_{SD}\left[N\right]}{P_{Star}\left[N\right]}\right] = \frac{2}{3} \tag{10}$$

In fact, the recurrence is almost the same for $N \geq 3$, we just need to add a symmetry multiplier:

$$P_{SD}\left[N+1\right] = \left(\frac{2}{N\left(N-1\right)}\right) + P_{SD}\left[N\right]\left(\frac{N-3}{N} + \frac{1}{2N}\right) \cdot \begin{cases} 1 & N \neq 4 \\ 2 & N = 4 \end{cases} \tag{11}$$

4.4 2-Hub star-like

The formula given in [KR2017] is mostly correct, we just need to add the same symmetry indicators as above. We will change m,n to s,t so we don't mix n with N:

$$\begin{aligned} P_{2H}\left[s,t\right] &= +a \cdot b \cdot \left(\tfrac{s-1}{(s+t+1)} + \tfrac{1}{(s+t+1)(t+1)}\right) P_{2H}\left[s-1,t\right] \\ &+a \cdot c \cdot \left(\tfrac{t-1}{(s+t+1)} + \tfrac{1}{(s+t+1)(s+1)}\right) P_{2H}\left[s,t-1\right] \end{aligned} \tag{12}$$

Where:

$$a = \begin{cases} 1 & s \neq t \\ \frac{1}{2} & s = t \end{cases} \tag{13}$$

$$b = \begin{cases} 1 & s-1 \neq t \\ 2 & s-1 = t \end{cases} \tag{14}$$

$$c = \begin{cases} 1 & s \neq t-1 \\ 2 & s = t-1 \end{cases} \tag{15}$$

As single-defect stars are just particular cases of the 2-Hub, we fill the borders with:

$$P_{2H}\left[s,1\right] = P_{SD}\left[s+3\right] \tag{16}$$

$$P_{2H}\left[1,t\right] = P_{SD}\left[t+3\right] \tag{17}$$

4.5 2-Hub by total leaves

We simply add over the diagonals:

$$P_{2H}\left[N\right] = \sum_{s=1}^{\left\lfloor \frac{N-2}{2} \right\rfloor} P_{2H}\left[s, N-s-2\right] \tag{18}$$

This seems to converge (more work needed!) to:

$$\lim_{n\to\infty}\left[\frac{P_{2H}\left[N\right]}{P_{Star}\left[N\right]}\right] = \frac{3}{2} \tag{19}$$

Which is reasonable looking at the spirals on the PDFs.

Something similar (perhaps the fractions on the limits increase to $\frac{4}{3}, \frac{5}{4}, \ldots$?) might happen with an increasing number of nucleus.

5 Transition/growth rules for the complete redirection model

5.1 Values

Looking at how each configuration generates every possible newt one and counting leaves, hubs and cores, we get the following changes (Figure 11):

{{2, 0, 0}, 1}								
{{2, 1, 0}, 1}								
{{2, 1, 0}, 1}								
{{3, 1, 0}, $\frac{2}{3}$}	{{2, 2, 0}, $\frac{1}{3}$}							
{{3, 1, 0}, $\frac{2}{3}$}		{{2, 2, 0}, $\frac{1}{3}$}						
{{4, 1, 0}, $\frac{1}{2}$}	{{3, 2, 0}, $\frac{1}{6}$}	{{3, 2, 0}, $\frac{1}{4}$}	{{2, 2, 1}, $\frac{1}{12}$}					
{{4, 1, 0}, $\frac{1}{2}$}		{{3, 2, 0}, $\frac{5}{12}$}			{{2, 2, 1}, $\frac{1}{12}$}			
{{5, 1, 0}, $\frac{2}{5}$}	{{4, 2, 0}, $\frac{1}{10}$}	{{4, 2, 0}, $\frac{23}{72}$}	{{3, 2, 1}, $\frac{1}{24}$}	{{3, 3, 0}, $\frac{1}{18}$}	{{3, 2, 1}, $\frac{1}{20}$}	{{2, 2, 2}, $\frac{1}{60}$}	{{3, 3, 0}, $\frac{1}{60}$}	

Figure 11: Counts Changes

Calculating the differences and collecting them (Figure 12):

	{1., 0., 0.}	{0., 1., 0.}	{0., 0., 1.}	{1., 1., -1.}
		1.		
	0.666667	0.333333		
	0.75	0.166667	0.0833333	
	0.769444	0.155556	0.0583333	0.0166667
	0.789468	0.140085	0.0521605	0.018287
	0.804241	0.130109	0.0470826	0.0185681
	0.816085	0.122309	0.043503	0.0181035
	0.825798	0.115999	0.0407486	0.017454
	0.833958	0.110721	0.0385453	0.0167754
	0.840944	0.106204	0.0367259	0.0161264
	0.847017	0.102271	0.0351869	0.0155257
	0.852364	0.0987996	0.0338601	0.0149766
	0.857122	0.0957021	0.0326988	0.0144767

Figure 12: Grouping the Differences

5.2 Rules

Since we'll be dealing with averages, we should consider two averages:

$$E_H = \frac{L + 2HH + HC}{H} \tag{20}$$

And:

$$E_C = \frac{HC + 2CC}{C} \tag{21}$$

Which are, respectively, the average number of edges per hub and the average number of edges per core (Table 1).

Provisional	Prob	Final	Cond Prob	Hub with 1 Leaf	L	H	C
L	$\frac{L}{N}$	H	1		+1	0	0
H	$\frac{H}{N}$	L	$\frac{L}{H \cdot E_H}$	False	0	+1	0
H		L	$\frac{L}{H \cdot E_H}$	True	0	0	+1
H		H	$\frac{2HH}{H \cdot E_H}$		+1	0	0
H		C	$\frac{HC}{H \cdot E_H}$		+1	+1	-1
C	$\frac{C}{N}$	H	$\frac{HC}{C \cdot E_C}$		+1	0	0
C		C	$\frac{2CC}{C \cdot E_C}$		+1	+1	-1

Table 1: Uncollected rules

Collecting the rules by their effect (Table 2):

L	H	C	Probability
+1	0	0	$\frac{L}{N} + \frac{H}{N} \cdot \frac{2HH}{H \cdot E_H} + \frac{C}{N} \cdot \frac{HC}{C \cdot E_C}$
0	+1	0	$\frac{H}{N} \cdot \frac{L}{H \cdot E_H} \cdot H_{Leafy}$
0	0	+1	$\frac{H}{N} \cdot \frac{L}{H \cdot E_H} \cdot H_{Single}$
+1	+1	-1	$\frac{H}{N} \cdot \frac{HC}{H \cdot E_H} + \frac{C}{N} \cdot \frac{2CC}{C \cdot E_C}$

Table 2: Collected rules

From this we can see how hard it is for a core of cores to be intact - they will break down when a new leaf appears and a new hub breaks down the group.

From the data (small N), it seems that:

$$\lim_{n \to \infty} \left[\frac{H_{Leafy}}{H_{Single}} \right] = 3 \tag{22}$$

And in order to roughly match the decay described in [KR2017] we use the following approximations (Table 3):

L	H	C	Probability
+1	0	0	$\frac{L}{N} + \frac{H}{N} \cdot \frac{1}{5} + \frac{C}{N} \cdot 1$
0	+1	0	$\frac{H}{N} \cdot \frac{3}{5}$
0	0	+1	$\frac{H}{N} \cdot \frac{1}{5}$

Table 3: Approximated rules

Implied in this approximation is that, on average, most of the edges of hubs come from leaves (reasonable) and most of the edges of cores from hubs (also reasonable, as explained above).

5.3 Iterations

For the progress of each relative frequency $\left(\frac{L}{N}, \frac{H}{N}, \frac{C}{N}\right)$ we use the update formula:

$$P_{N+1} = \frac{N \cdot P_N + Probability}{N+1} \tag{23}$$

For N up to 5000 we see how leaves quickly dominate and the cores rise and decay (Figure 13):

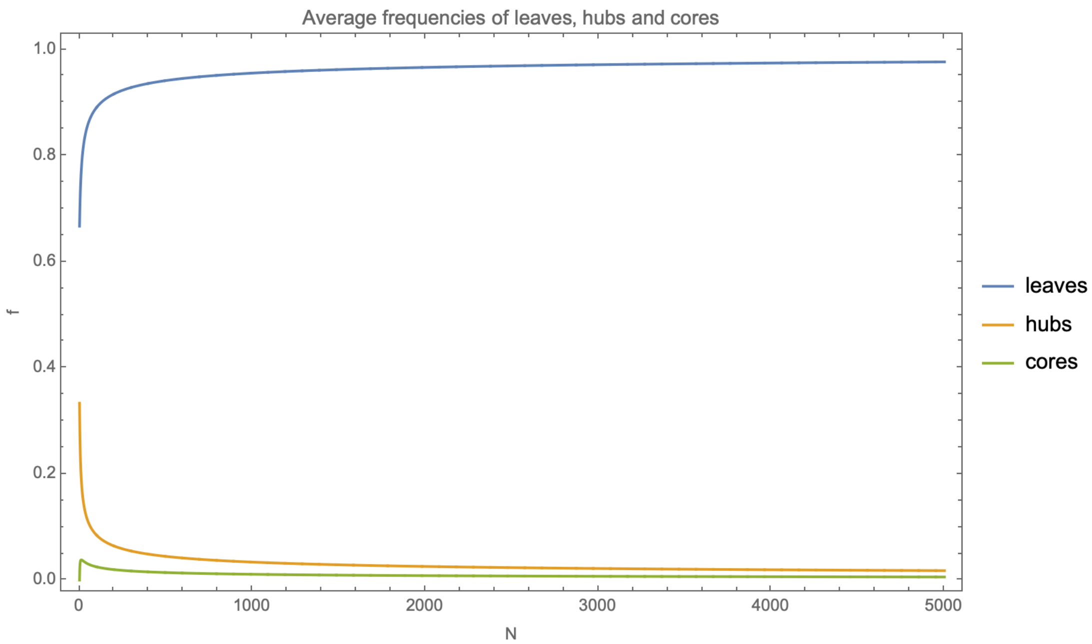

Figure 13: Elements frequency over 5000 iterations

Comparing the nucleus frequency for N up to 5000 with the decay from [KR2017] (Figure 14):

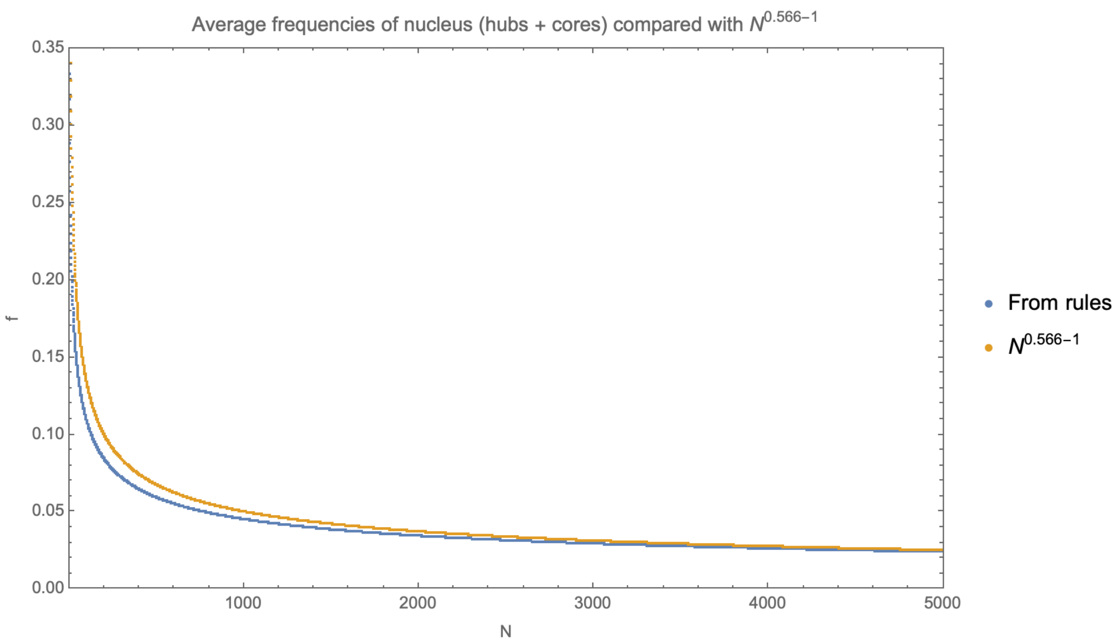

Figure 14: Nucleus frequency over 5000 iterations

We can switch to Logs for N up to 50000 (Figure 15):

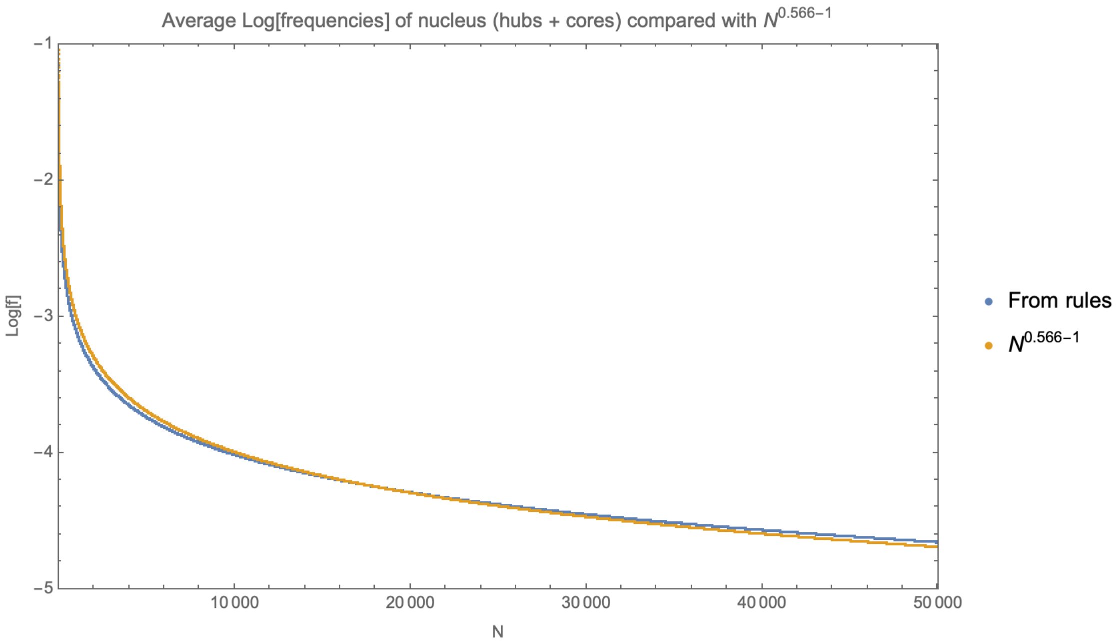

Figure 15: Nucleus Log[frequency] over 50000 iterations

Not as good but not that bad for N up to 500000, even without changing the parameters of the model (Figure 16):

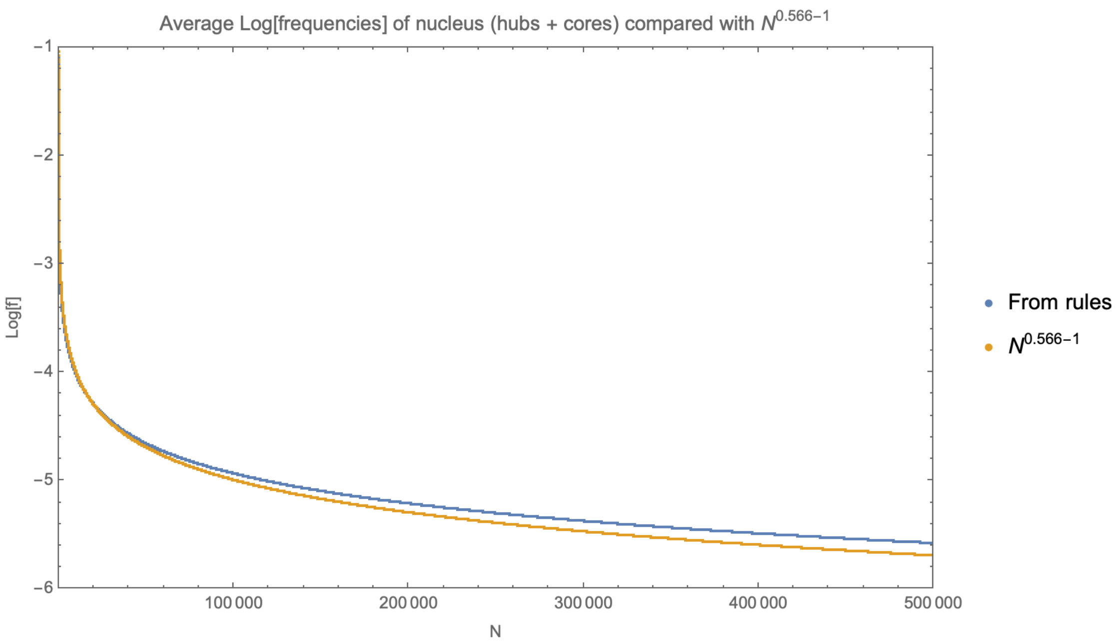

Figure 16: Nucleus Log[frequency] over 500000 iterations

We could calibrate the model better; it's not perfect for small N (Figure 17), but we wanted to get a better behavior for larger values of N:

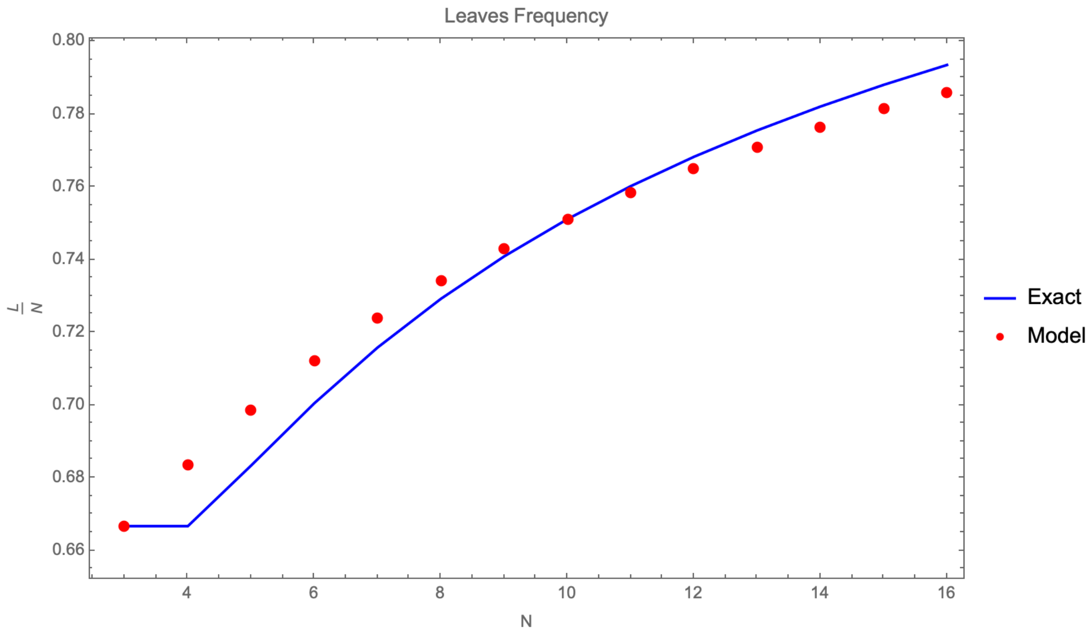

Figure 17: Leaves frequency for small N

6 Partitions and Rules

6.1 Partitions

The set of possible elements of the distribution of degrees of the nucleus is equivalent to the collection of restricted (no 1s) partitions of $2(N-1)-L$ with $N-L$ elements, with L from 2 to $N-1$. An example for 10 nodes is shown on Table 4 (so many 2s!):

L	Partitions
9	$\{\{9\}\}$
8	$\{\{8,2\},\{7,3\},\{6,4\},\{5,5\}\}$
7	$\{\{7,2,2\},\{6,3,2\},\{5,4,2\},\{5,3,3\},\{4,4,3\}\}$
6	$\{\{6,2,2,2\},\{5,3,2,2\},\{4,4,2,2\},\{4,3,3,2\},\{3,3,3,3\}\}$
5	$\{\{5,2,2,2,2\},\{4,3,2,2,2\},\{3,3,3,2,2\}\}$
4	$\{\{4,2,2,2,2,2\},\{3,3,2,2,2,2\}\}$
3	$\{\{3,2,2,2,2,2,2\}\}$
2	$\{\{2,2,2,2,2,2,2,2\}\}$

Table 4: Restricted partitions for N=10

As discussed above, all the different rules for choosing the attachment node end up generating the same networks, just with different frequencies. We can present the networks sorted by their frequency for N=7 for comparison (Figures 18, 19, 20 and 21):

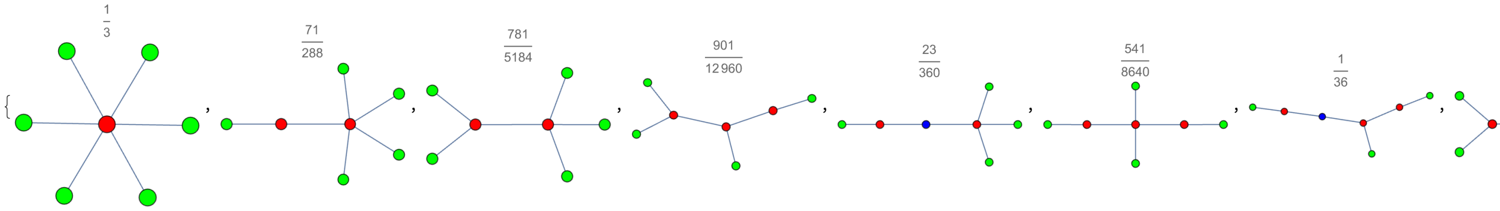

Figure 18: Enumeration of all networks configurations for N=7 (Complete Redirection)

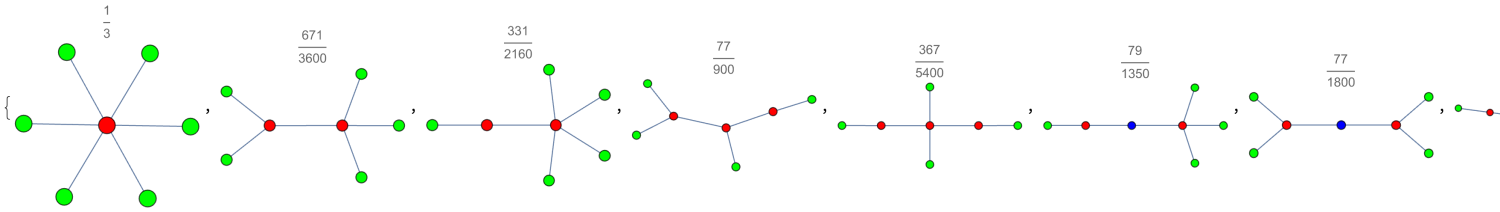

Figure 19: Enumeration of all networks configurations for N=7 (Complete Fuzzy Redirection)

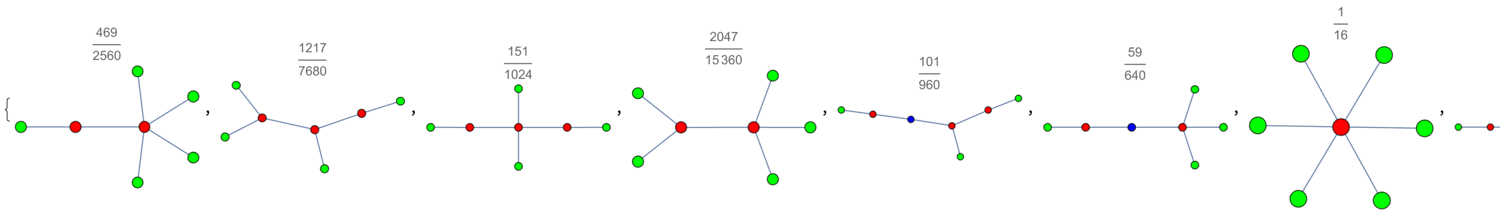

Figure 20: Enumeration of all networks configurations for N=7 (Partial Redirection $pr = 0.50$)

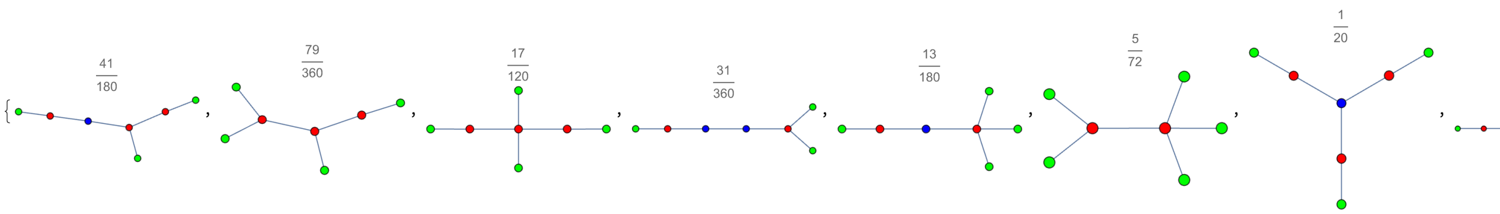

Figure 21: Enumeration of all networks configurations for N=7 (No Redirection)

Now we start to understand why separating cores from hubs is useful. Cores are quite rare, and their maximum degree is not $N-1$ like a hub, but $\left\lfloor \frac{N-1}{2} \right\rfloor$ (we can see this star-like configuration as the next-to-last configuration in the first 3 figures above). Even the “snake”, the configuration with the maximum amount of cores, is quite rare in all models (and with so many cores their degree is 2). So lots of 2s.

6.2 Rules

Let’s go back to our rules tables and write down the other models (Tables 5, 6 and 7):

Prov	Prob	Final	Cond Prob	$D_{H_j} = 2$	L	H	C	$PDF\left(\Delta D_H\right)$	$PDF\left(\Delta D_C\right)$
L	$\frac{L}{N}$	H	1		+1	0	0	$\mathbb{P}\left[D_{H_{jj}} \uparrow\right] = \frac{D_{H_j}}{\sum D_H}$	0
H	$\frac{H}{N}$	L	$\frac{L}{H \cdot E_H}$	False	0	+1	0	New $D_H = 2$	0
H		L	$\frac{L}{H \cdot E_H}$	True	0	0	+1	$f_1\left(geometry\right) \downarrow$	$f_1\left(geometry\right) \uparrow$
H		H	$\frac{2HH}{H \cdot E_H}$		+1	0	0	$f_2\left(geometry\right) \uparrow$	0
H		C	$\frac{HC}{H \cdot E_H}$		+1	+1	-1	$f_3\left(geometry\right) \uparrow$	$f_3\left(geometry\right) \downarrow$
C	$\frac{C}{N}$	H	$\frac{HC}{C \cdot E_C}$		+1	0	0	$f_4\left(geometry\right) \uparrow$	0
C		C	$\frac{2CC}{C \cdot E_C}$		+1	+1	-1	$f_5\left(geometry\right) \uparrow$	$f_5\left(geometry\right) \downarrow$

Table 5: Uncollected rules (Complete Redirection)

Prov	Prob	Final	Cond Prob	$D_{H_j} = 2$	L	H	C	$PDF\left(\Delta D_H\right)$	$PDF\left(\Delta D_C\right)$
L	$\frac{L}{N}$	H	1		+1	0	0	$\mathbb{P}\left[D_{H_{jj}} \uparrow\right] = \frac{1}{H}$	0
H	$\frac{H}{N}$	L	$\frac{L}{H \cdot E_H}$		0	+1	0	$f_1\left(geometry\right) \uparrow$	0
H		L	$\frac{L}{H \cdot E_H}$		0	0	+1	$\left(1 - f_1\left(geometry\right)\right) \downarrow$	$\left(1 - f_1\left(geometry\right)\right) \uparrow$
H		H	$\frac{2HH}{H \cdot E_H}$		+1	0	0	$\mathbb{P}\left[D_{H_{jj}} \uparrow\right] = \frac{1}{H-1}$	0
H		C	$\frac{HC}{H \cdot E_H}$		+1	+1	-1	$f_2\left(geometry\right) \uparrow$	$f_2\left(geometry\right) \downarrow$
C	$\frac{C}{N}$	H	$\frac{HC}{C \cdot E_C}$		+1	0	0	$\mathbb{P}\left[D_{H_{jj}} \uparrow\right] = \frac{1}{H}$	0
C		C	$\frac{2CC}{C \cdot E_C}$		+1	+1	-1	$f_3\left(geometry\right) \uparrow$	$f_3\left(geometry\right) \downarrow$

Table 6: Uncollected rules (Complete Fuzzy Redirection)

Prov	Prob	Final	Cond Prob	$D_{H_j} = 2$	L	H	C	$PDF\left(\Delta D_H\right)$	$PDF\left(\Delta D_C\right)$
	$\frac{H}{N}$	H	1		+1	0	0	$\mathbb{P}\left[D_{H_{jj}} \uparrow\right] = \frac{1}{H}$	0
	$\frac{L}{N}$	L	$\frac{L}{H \cdot E_H}$	False	0	+1	0	New $D_H = 2$	0
		L	$\frac{L}{H \cdot E_H}$	True	0	0	+1	$f_1\left(geometry\right) \downarrow$	$f_1\left(geometry\right) \uparrow$
	$\frac{C}{N}$	C	$\frac{2CC}{C \cdot E_C}$		+1	+1	-1	$f_2\left(geometry\right) \uparrow$	$f_2\left(geometry\right) \downarrow$

Table 7: Uncollected rules (No Redirection)

The main change between the Complete Redirection and the Complete Fuzzy Redirection is that the PDF of the degrees of the existing hubs is not used anymore; other geometry dependences on the changes of the degrees of hubs are also replaced by uniform distributions. But the change in the first row is very important; there's no built-in feedback to stretch this PDF to a barbell.

And the last set of rules imply a cross-dependency that will keep proportions balanced.

7 Drift

Let's look at the PDFs for the No Redirection model first (Figure 22):

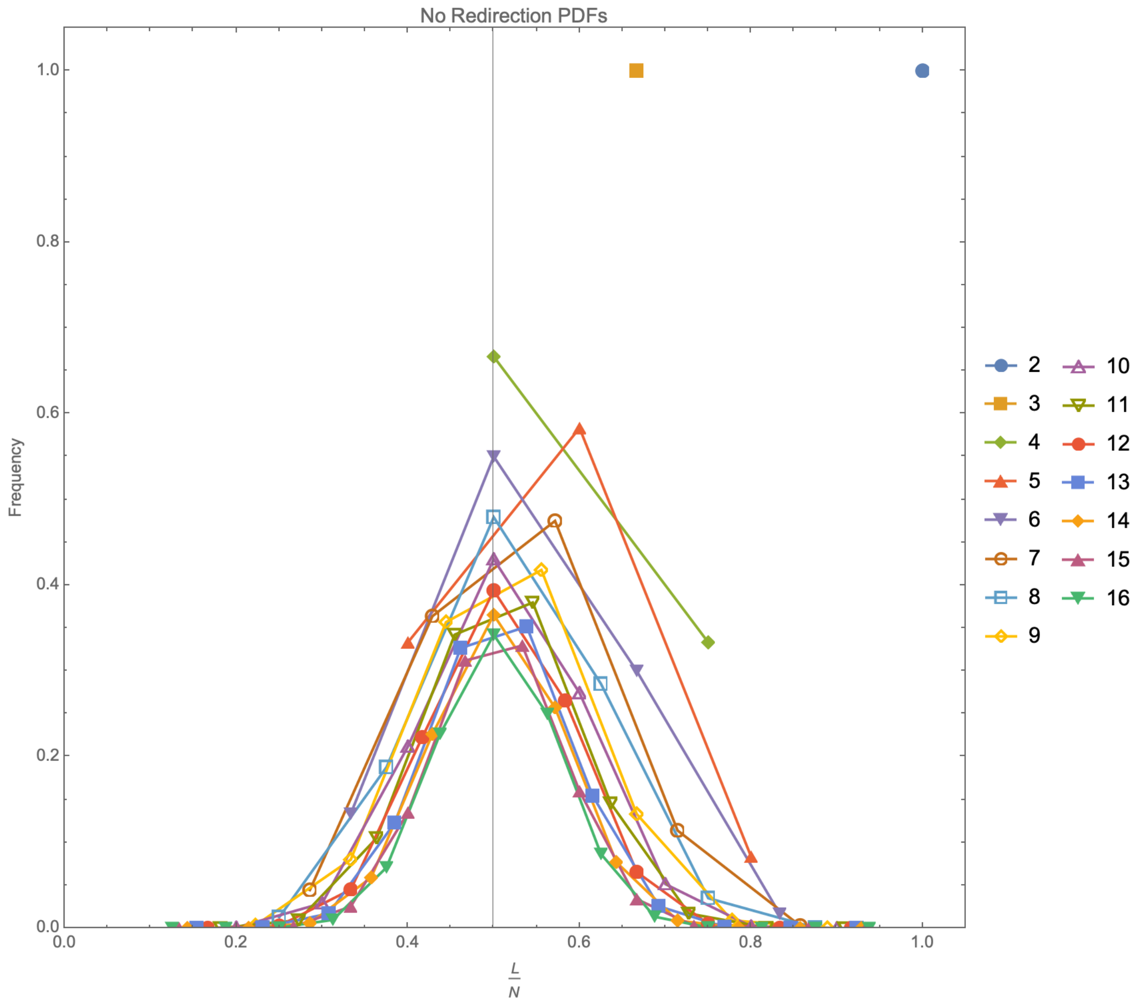

Figure 22: PDFs for $\frac{L}{N}$ up to N=16 - No Redirection

The PDFs get narrower while centered around 0.5. There's some pressure from the borders to the center keeping the network balanced between leaves and (hubs + cores).

We can plot the transition probabilities for $\frac{Leaves}{Nodes}$ below:

$$\mathbb{P}\left[\frac{L+1}{N+1}\right] = \mathbb{P}\left[\frac{L}{N}\right] \cdot \mathbb{P}\left[\frac{L}{N} \to \frac{L+1}{N+1}\right] + \mathbb{P}\left[\frac{L+1}{N}\right] \cdot \mathbb{P}\left[\frac{L+1}{N} \to \frac{L+1}{N+1}\right] \tag{24}$$

Which can be seen as:

$$\mathbb{P}\left[\frac{L+1}{N+1}\right] = \mathbb{P}\left[\frac{L}{N}\right] \cdot \mathbb{P}\left[\frac{L}{N} \to \frac{L+1}{N+1}\right] + \mathbb{P}\left[\frac{L+1}{N}\right] \cdot \left(1 - \mathbb{P}\left[\frac{L+1}{N} \to \frac{L+2}{N+1}\right]\right) \tag{25}$$

With the boundary:

$$\mathbb{P}\left[\frac{N}{N+1}\right] = \mathbb{P}\left[\frac{N-1}{N}\right] \cdot \mathbb{P}\left[\frac{N-1}{N} \to \frac{N}{N+1}\right] \tag{26}$$

So we go backwards, using the last equation to calculate:

$$\mathbb{P}\left[\frac{N-1}{N} \to \frac{N}{N+1}\right] = \frac{N-1}{N} \tag{27}$$

And then the recurrence:

$$\mathbb{P}\left[\frac{N-k}{N} \to \frac{N-k+1}{N+1}\right] = \frac{\mathbb{P}\left[\frac{N-k+1}{N+1}\right]}{\mathbb{P}\left[\frac{N-k+1}{N}\right]} - \frac{\mathbb{P}\left[\frac{N-k}{N}\right]}{\mathbb{P}\left[\frac{N-k+1}{N}\right]}\left(1 - \mathbb{P}\left[\frac{N-k+1}{N} \to \frac{N-k+2}{N+1}\right]\right) \tag{28}$$

Will give us all the transition probabilities to the "side" that increases the relative frequency of leaves.

We can show the difference between these probabilities and 0.5 as a "drift" towards leafiness:

$$D\left[\frac{N-k}{N} \to \frac{N-k+1}{N+1}\right] = \mathbb{P}\left[\frac{N-k}{N} \to \frac{N-k+1}{N+1}\right] - \frac{1}{2} \tag{29}$$

And plot this over n (Figure 23):

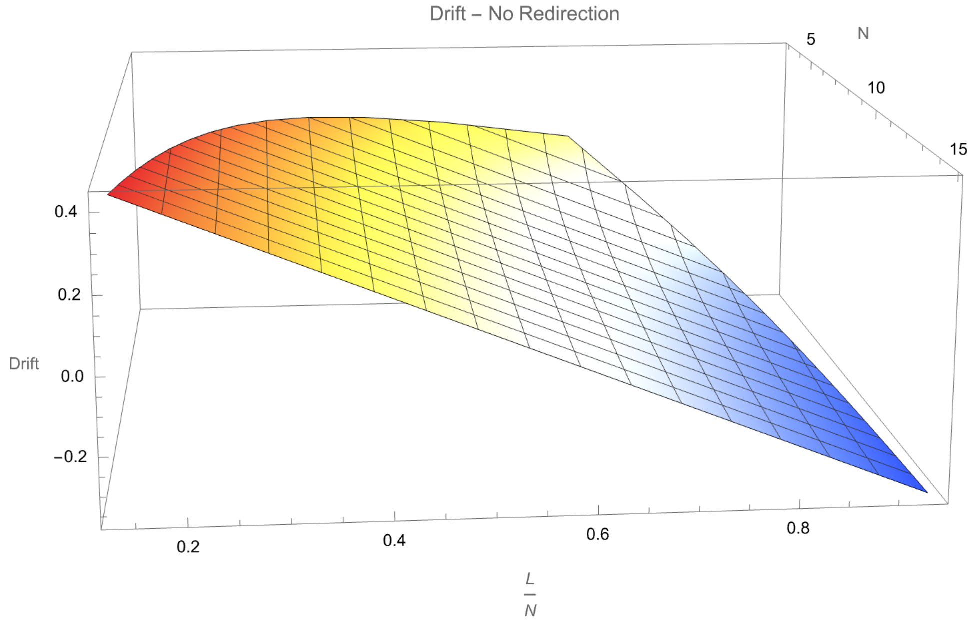

Figure 23: Drift - No Redirection

We can see that the transition probabilities force the PDF back to the centre, as the leafy half has a negative drift and the non-leafy part has a positive drift.

Revisiting the PDF for the Complete Redirection model (Figure 24):

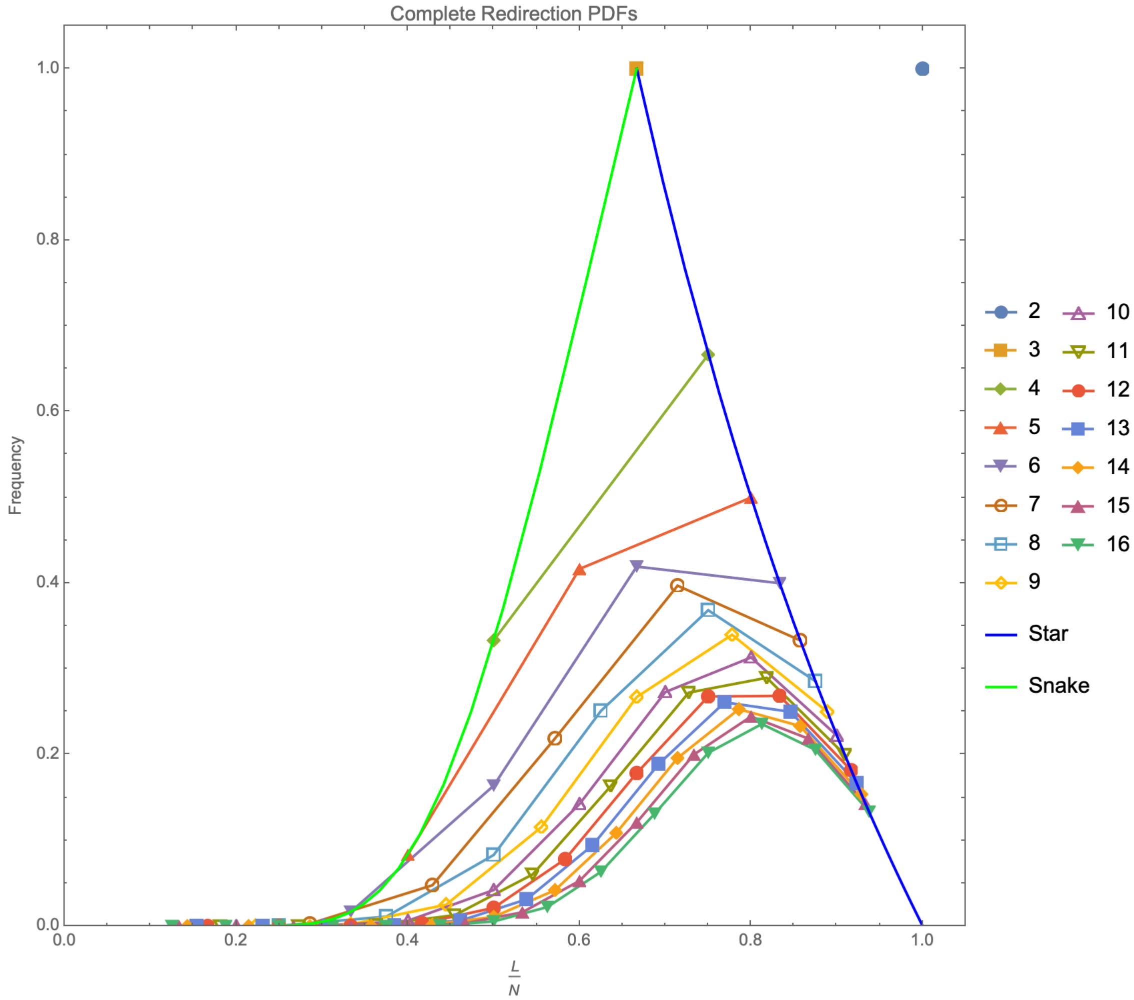

Figure 24: PDFs for $\frac{L}{N}$ up to N=16 - Complete Redirection

We can see the PDF shifting to the right (remember that in the limit, the ratio between the frequencies of N-2 leaves and N-1 leaves is $\frac{3}{2}$). The drifts look remarkably symmetrical (Figure 25); in fact, both borders are equal to $\frac{N-1}{N}$. And the minimum for each N seems to increase with N, even if slightly so.

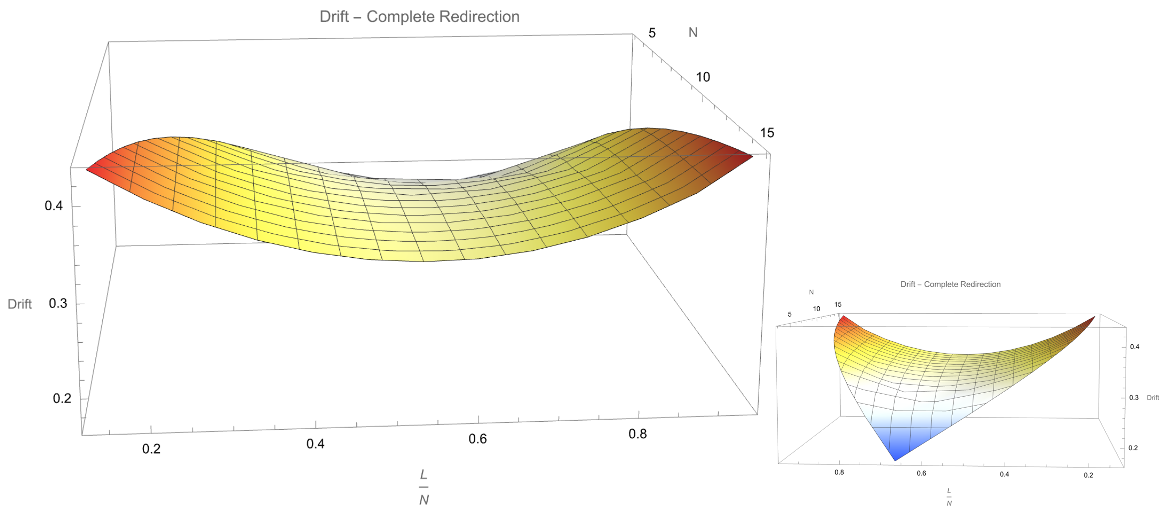

Figure 25: Drift - Complete Redirection

Let's compare this with the Complete Fuzzy Redirection. The PDFs and drifts are below (Figures 26 and 27):

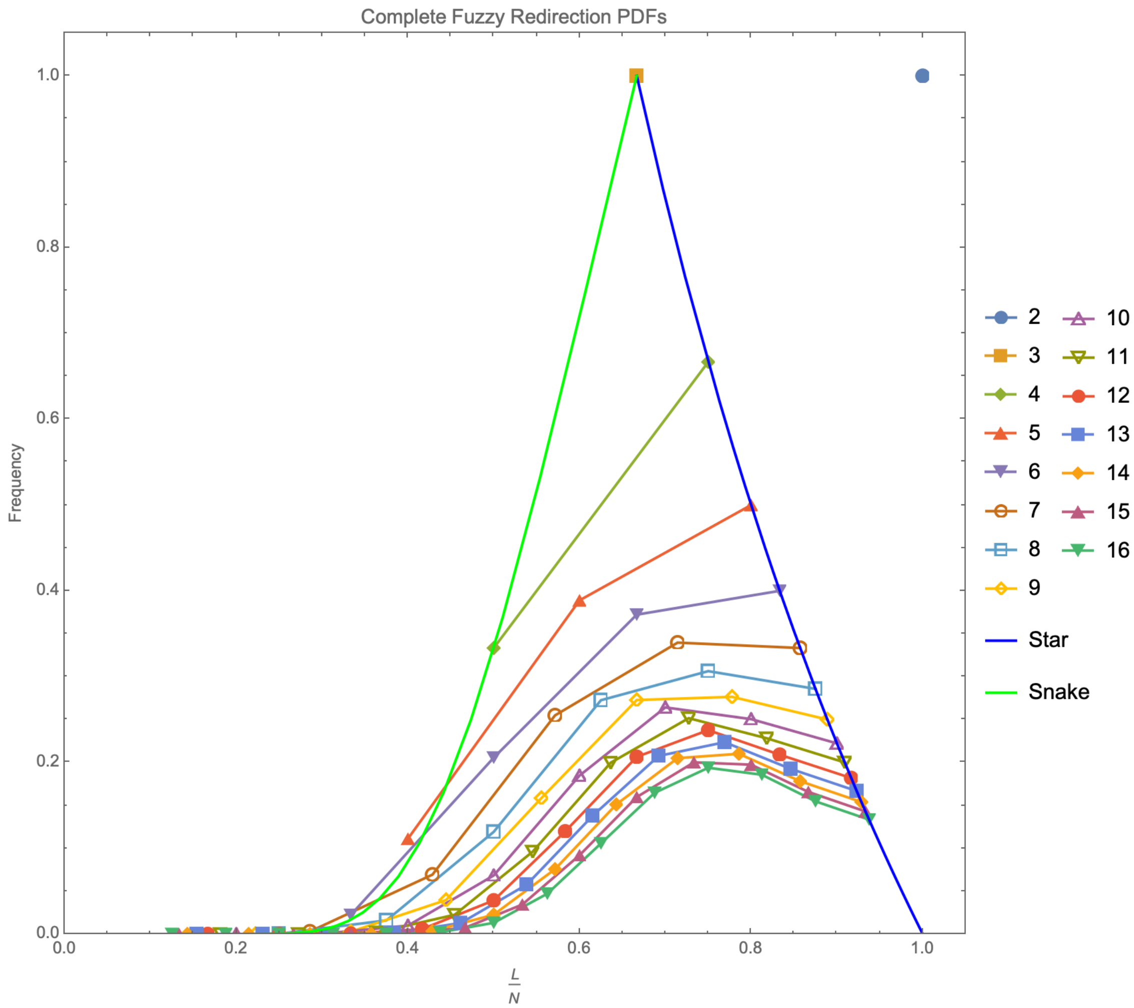

Figure 26: PDFs for $\frac{L}{N}$ up to N=16 - Complete Fuzzy Redirection

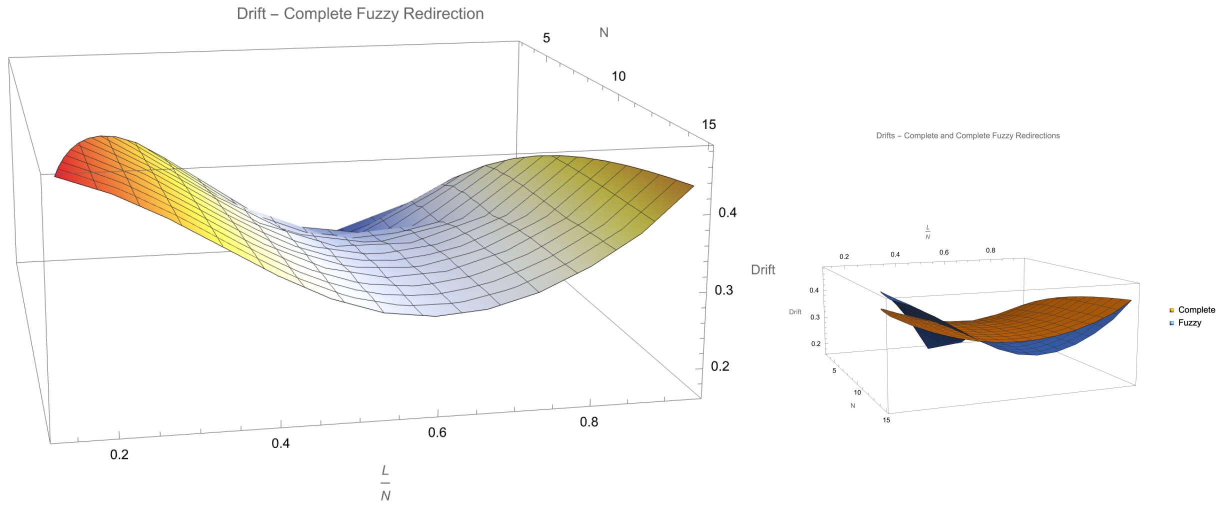

Figure 27: Drift - Complete Fuzzy Redirection

The drifts are weaker in most of the region; the results in the PDFs can be seen clearly in Figure 28; although the values for $\frac{N-1}{N}$ are the same, the "wave" doesn't rise so high:

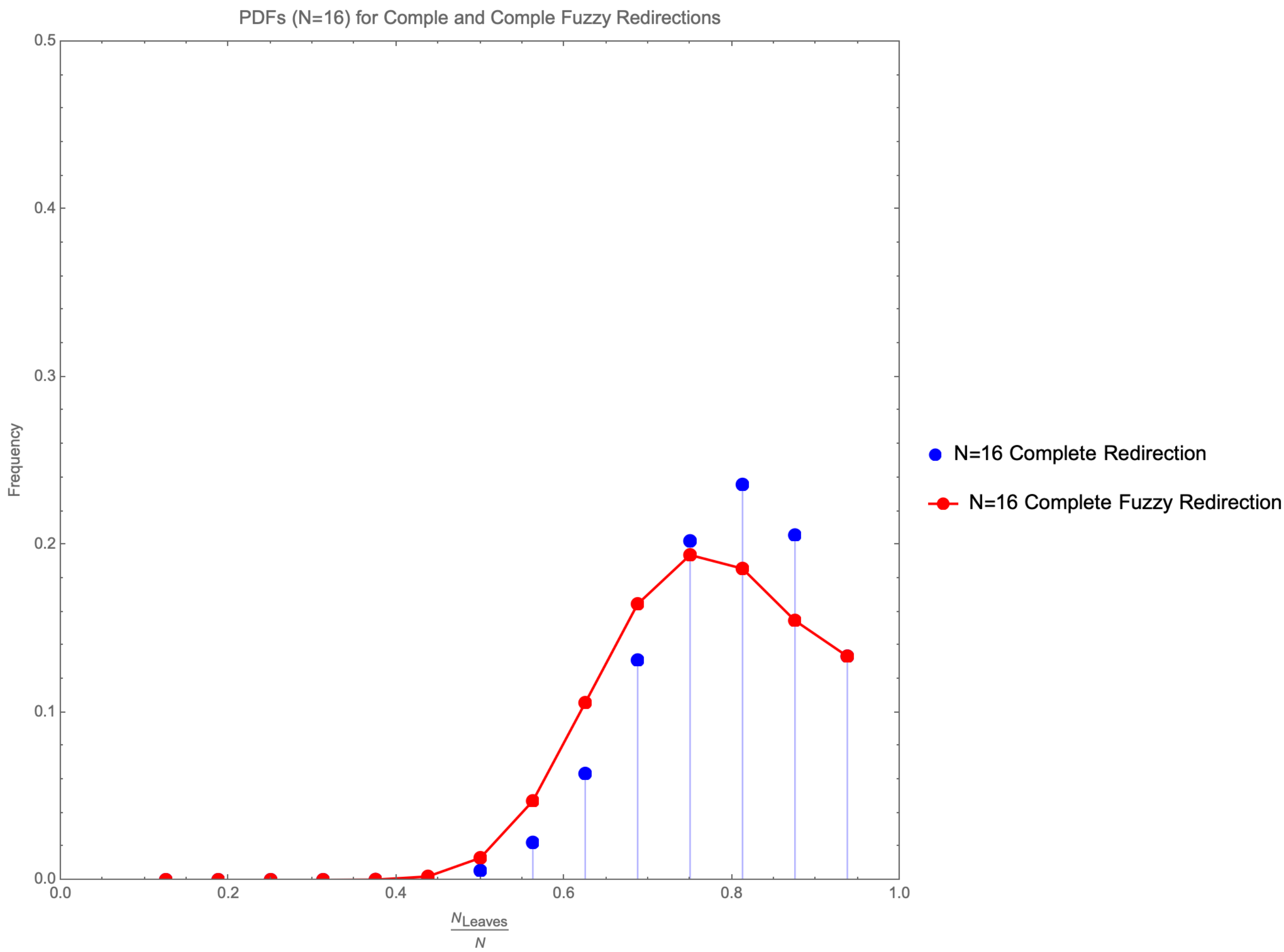

Figure 28: PDFs for N=16 - Complete and Complete Fuzzy Redirections

Just a positive drift all around is not enough to get us to zero mass in the nucleus, as Figure 29 shows for the partial Redirection with pr=0.75:

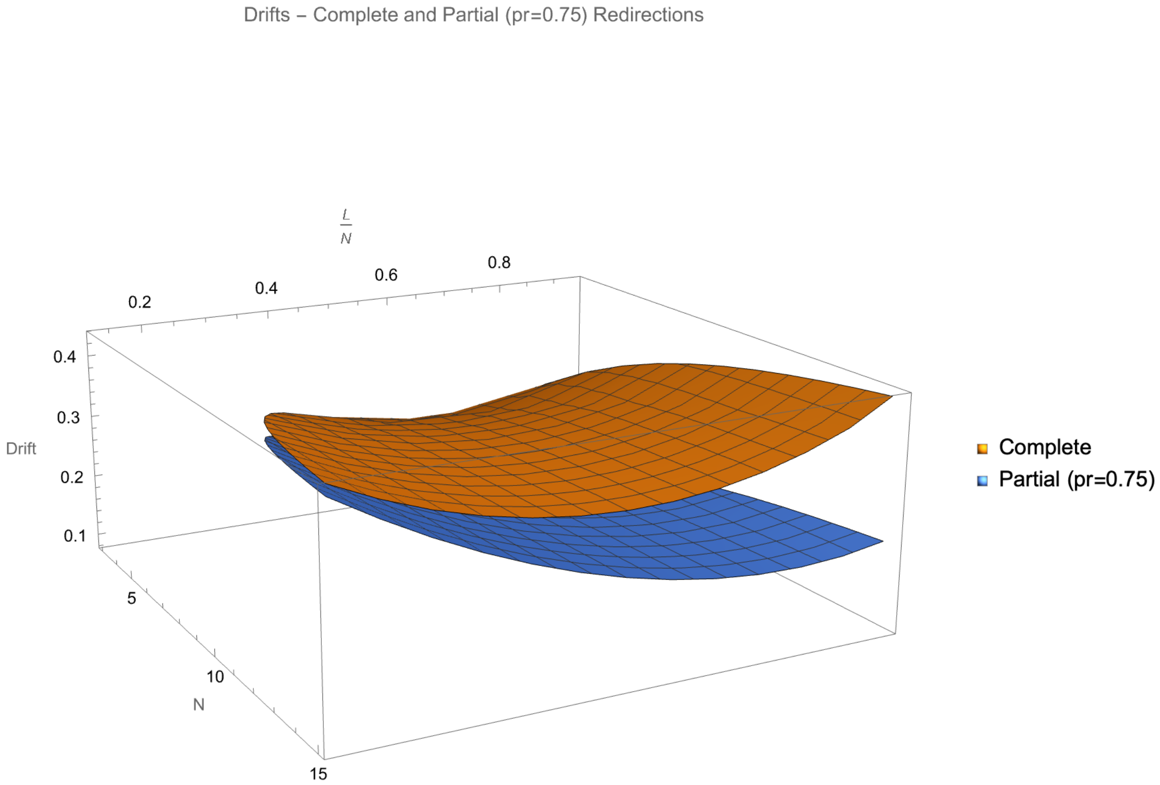

Figure 29: Drift - Partial Redirection (pr=0.75)

8 Question 1: Is there a redirection probability lower than 1 that leads to the same kind of sublinear growth for the nucleus?

The answer is No. Let's look at the difference between drifts (Complete - Partial) for redirection probabilities pr=0.99 and pr=0.9999 (Figure 30):

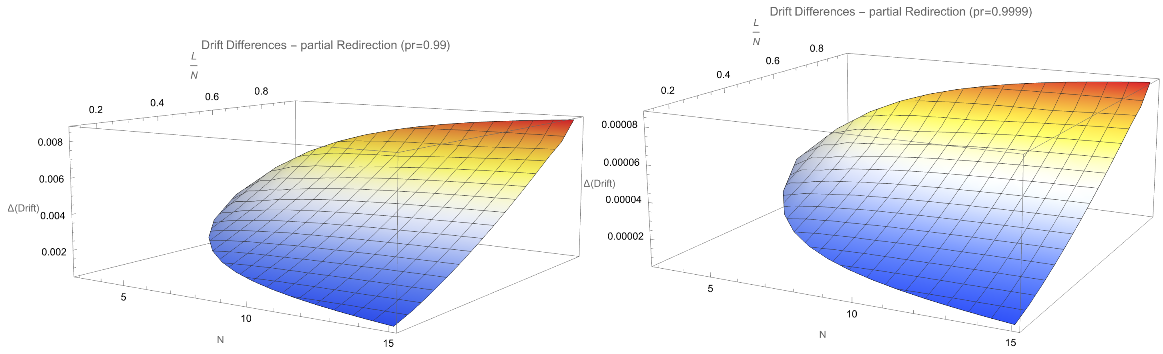

Figure 30: Drift Differences - Partial Redirections

If the mass of leaves went to 1, it would be generating hubs at a $(1 - pr)$ pace; this is the "headwind" seen in Figure 30: the higher the frequency of leaves, the lower the drift becomes.

For really low values of $(1 - pr)$, we cannot trust simulations too much.

9 Question 2: Is the geometry component necessary for the nucleus fraction to vanish?

The answer seems to be Yes. Although the drifts difference goes to zero as the frequency increases (Figure 31), Figure 3 hints at a flattening of the relative frequency of leaves that can possibly be explained by integrating the drifts (or their difference). It seems that the drifts from the Complete Redirection network are just enough to make the nucleus mass disappear.

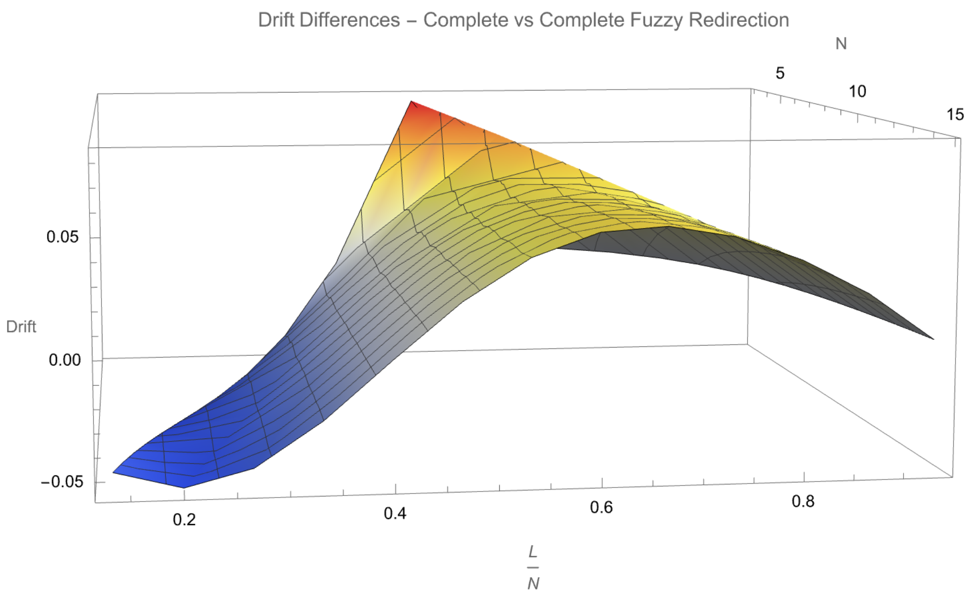

Figure 31: Drift Differences - Complete vs Complete Fuzzy Redirections

Integrating the drift might help to determine μ from the [KR2017] paper.

10 Conclusions and further work

With the automated enumeration up to N=16 we confirmed, corrected and derived closed formulas some some of the probabilities in this model, and by contrasting the behavior of hubs and cores we created simple rules that, without changing the parameters over time, exhibit roughly the same relative frequencies as described in [KR2017]. We can improve the model with a term structure of parameters (adapting it to the actual frequencies). Because the model looks only at the averages, it doesn't necessarily fully explain the behavior for the distribution of the connectivity within the nucleus, but the PDFs generated by the full enumeration might be helpful in that.

We look further at the other possible redirection configurations, and try to answer whether only the complete redirection (with its barbell-like feedback on existing hubs) is the only possible configuration where the nucleus mass goes to zero.

A companion Mathematica Notebook is available (currently not as well organized as it should). The simulation code can be improved, and the small N results might be extended (and hopefully parametrized as a function of n and/or a recurrence).

The fuzzy model is interesting, but it is also the slowest to simulate (at every iteration we must map leaves, hubs and cores).

Thanks to Steven Strogatz for drawing attention to the paper and for the comments, and to Sid Redner and Yuanzhao Zhang for analysis and comments as well.

References

[KR2017] Krapivsky, P. L. and Redner, S., Emergent Network Modularity, 2017

[C2022] Carreira, M. C. S., Exact values for small N and a simple recursion for the complete redirection network, 2022