Analysis of guifi.net's Topology

Extension of Results

Llorenç Cerdà-Alabern
Universitat Politècnica de Catalunya, Computer Architecture Dep.
Jordi Girona 1-3, E-08034, Barcelona, Spain
llorenc@ac.upc.edu

May 4, 2012

***Abstract*—This report extends the analysis carried out in a previous work [1] about the topology of guifi.net wireless community network. The main objective is validating the topology generator proposed in [1] for guifi.net like topologies by considering a larger number of zones. The numerical results obtained in this report are in line with those obtained before, confirming the topology generator. Additionally, new results are presented, as the link length distribution.**

I. Introduction

This report extends the analysis of guifi.net's topology carried out in [1]. More specifically, in [1] only two zones were considered. In this report we try to answer the questions: To what extent are the conclusions obtained in [1] characteristic of guifi.net? Is the topology generator proposed [1] valid for any zone of guifi.net, or the whole network itself? We do so by considering a larger set of guifi.net zones. The results presented in this report have been obtained using the CNML file downloaded from the following URL on April the 19th, 2012. See [1] for details about the CNML format.
http://guifi.net/en/guifi/cnml/3671/detail

The CNML file is organized in a hierarchy of zones. The root zone is named *guifi.net World*. Each zone can be divided in other zones that cover smaller geographical areas. We shall refer to each division as *levels*, being the root zone level 1. In this report we have considered the zones up to level 5. From them, we have formed the topology graphs as explained in [1], and we have chosen the 15 leaf zones having the graphs with the largest number of nodes. Figure 1 shows the zone tree that resulted following this approach. In the analysis we have also considered Catalunya level 4 zone. This allows studying whether the aggregated zones keep the same characteristics as the leafs. We have considered Catalunya because most of its leafs form a connected graph. At lower levels there are many disconnected clusters, which are connected through the Internet, but not using links reported in the CNML file. The zones analyzed in this report are shown in figure 1 numbered in decreasing order of their number of nodes (e.g. Osona is the leaf-zone up to level 5 having the largest number of nodes).

The rest of the report is organized as follows: In section II the same graphs obtained in [1] are built for the zones described above. In secion III these results are discussed and compared with those presented in [1]. Section IV analyzes the link length distribution. Finally, section V ends with some concluding remarks.

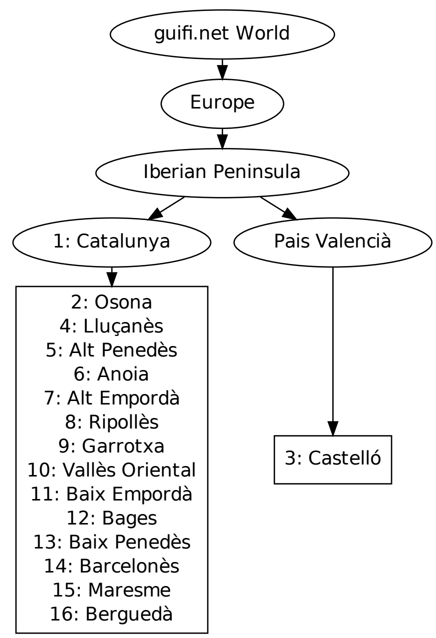


Figure 1. Zones considered in this report, ordered by the number of nodes of their base-graph.

II. Numerical Results

We have formed the *base-graph* and *core-graph* of each zone as explained in [1]. Recall that the core-graph is obtained by removing the terminal nodes of the base-graph. As in [1], the terminals of the base-graph are referred to as *hidden-terminals*.

Table I (a) summarizes the number of nodes, links, and degree of base-graph and core-graph. For the degree it is given the minimum, mean and maximum values. Table I (b) gives the values of the main parameters obtained for these graphs, namely: rank and hops exponents of the power law fitting of base and core graphs, and the parameters of the gamma distribution fitting of the number of hidden-terminals. For each fitting table I (b) gives the sample correlation coefficient (ρ). The last rows of table I (b) give the sample mean (μ),

zone	base-graph		core-graph	
	node/links	degree	node/links	degree
1	10,625/10,949	1/2.06/476	735/1,059	1/2.88/30
2	6,112/6,244	1/2.04/476	266/398	1/2.99/29
3	1,548/1,577	1/2.04/176	70/99	1/2.83/14
4	947/963	1/2.03/93	47/63	1/2.68/8
5	848/856	1/2.02/107	54/62	1/2.3/8
6	384/395	1/2.06/65	52/63	1/2.42/11
7	343/369	1/2.15/39	43/69	1/3.21/9
8	332/335	1/2.02/148	12/15	1/2.5/7
9	257/313	1/2.44/58	44/100	1/4.54/13
10	256/256	1/2/44	20/20	1/2/5
11	244/251	1/2.06/27	30/37	1/2.47/10
12	242/250	1/2.07/66	32/40	1/2.5/9
13	190/190	1/2/40	9/9	1/2/5
14	181/190	1/2.1/18	54/63	1/2.33/9
15	170/172	1/2.02/82	13/15	1/2.31/4
16	161/163	1/2.02/28	22/24	1/2.18/5

(a)

zone	base-graph		core-graph		hidden terminals			
	rank/ρ	hops/ρ	rank/ρ	hops/ρ	shape	rate	mean	ρ
1	-0.43/0.74	3.6/0.99	-0.61/0.95	2.13/1	0.21	0.015	13.5	1.00
2	-0.33/0.64	3.8/0.99	-0.62/0.95	2.23/1	0.25	0.011	22.0	0.99
3	-0.38/0.67	4.7/1.00	-0.63/0.94	1.77/1	0.40	0.019	21.1	0.99
4	-0.38/0.67	3.8/0.99	-0.58/0.91	1.43/1	0.38	0.020	19.1	0.96
5	-0.48/0.74	3.7/1.00	-0.63/0.94	1.54/1	1.43	0.097	14.7	0.99
6	-0.57/0.85	2.6/0.99	-0.63/0.95	1.57/1	0.32	0.050	6.4	0.98
7	-0.62/0.84	2.8/1.00	-0.64/0.93	1.62/1	0.90	0.129	7.0	0.95
8	-0.36/0.63	4.8/1.00	-0.74/0.90	1.22/1	0.47	0.018	26.7	0.98
9	-0.69/0.88	3.1/1.00	-0.82/0.93	1.71/1	0.27	0.056	4.8	0.98
10	-0.51/0.75	2.5/0.99	-0.64/0.93	1.04/1	0.91	0.077	11.8	0.97
11	-0.61/0.83	2.6/1.00	-0.71/0.96	1.47/1	1.33	0.186	7.1	0.95
12	-0.59/0.84	2.1/0.98	-0.62/0.93	1.42/1	1.38	0.210	6.6	0.98
13	-0.44/0.67	3.5/1.00	-0.78/0.96	0.97/1	0.95	0.047	20.1	0.93
14	-0.69/0.94	1.9/1.00	-0.54/0.91	1.33/1	0.80	0.340	2.4	0.99
15	-0.48/0.75	3.1/0.97	-0.45/0.84	1.02/1	0.37	0.031	12.1	0.99
16	-0.61/0.85	2.0/0.99	-0.58/0.91	1.12/1	0.50	0.079	6.3	0.99
μ	-0.52	3.1	-0.64	1.4	0.68	0.087	13	0.98
c_i	0.067	0.51	0.052	0.19	0.23	0.048	3.9	0.01
σ	0.12	0.91	0.094	0.34	0.42	0.091	7.3	0.019
c_v	-0.23	0.29	-0.15	0.23	0.63	1	0.58	0.019

(b)

Table I

SUMMARY OF ZONE GRAPHS. FOR NODE DEGREE IT IS GIVEN MIN/MEAN/MAX. FOR THE RANK AND HOPS IT IS GIVEN EXPONENT/SAMPLE CORRELATION COEFFICIENT (ρ).

its 95% confidence interval (c_i), the standard deviation (σ) and the coefficient of variation ($c_v = \sigma/\mu$) of the values corresponding to the leaf zones (all but 1: Catalunya) in the respective columns.

We have first reproduced the main graphs obtained in [1] for Catalunya zone. More specifically:

- Figure 4 shows the base and core graphs.
- Figure 5 shows the rank $\log_{10}$-$\log_{10}$ plots of base and core graphs.
- Figure 6 shows the hops count $\log_{10}$-$\log_{10}$ plots of base and core graphs.
- Figure 8 shows: (a) the hidden-terminals CECDF semi-$\log_{10}$ plots; (b) the scatter plots of hidden-terminals vs core-degree; and (c) the average number of hidden-terminals vs core-degree.

Additionally, figure 7 shows the link length distribution of Catalunya zone. This analysis was not carried out in [1]. However, link length distribution might be of interest in a wireless network, due to its strong influence on the signal transmission in the radio channel. Link length is analyzed in section IV.

In figure 4 the nodes are placed in their geographical position, and the axes give the distances in km with respect of the origin of coordinates. The boxes shown in the figure correspond to the Catalunya leaf zones listed in figure 1. Figure 4 shows that zones 5 and 13 are empty. This is because in the CNML file there were not reported links between these zones and the others. Therefore, they were not included in the graph, in order to keep it connected.

For the leaf zones listed in figure 1, figures 9 and 10 show their base and core graphs, respectively; figures 12 and 13 show the rank $\log_{10}$-$\log_{10}$ plots of base and core graphs, respectively; figure 14 shows the hidden-terminals CECDF semi-$\log_{10}$ plots. Scatter plots of hidden-terminals and their averages are not given because most of the leaf zones have not enough points to produce representative plots. Finally, the link length distributions of the leaf zones are given in figure 15.

III. DISCUSSION

As shown in table I(a), the number of nodes of the zones analyzed varies more than an order of magnitude: from 161 to 6,112. Additionally, table I(b) shows that base-graph of all zones, except zone 14 (Barcelonés), are not well fitted by a power-law. In fact, the correlation coefficient (ρ) is in the range $[0.63, 0.88]$ for these zones, while it is 0.94 for Barcelonés. The rank plots shown in figure 12 depict these deviations from the power law fitting. We note that all zones, but Barcelonés, correspond to rural areas. Therefore, this result confirms the topology differences described in [1] between rural and urban graphs of guifi.net.

Regarding the core-graph, table I(b) shows that it is indeed much better fitted by a power-law: with the exception of zone 15, which has $\rho = 0.84$, ρ is now in the range $[0.90, 0.96]$. Nevertheless, even for zone 15 the power law fitting improves in the core with respect the base graph. Additionally, zone 15 has a small number of nodes (170), and thus, cannot be considered very representative. Figures 12

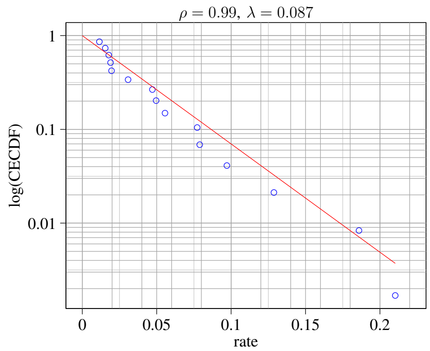


Figure 2. Semi-log plot of the CECDF of the rate parameter of the gamma distribution of hidden-terminals of the leaf zones.

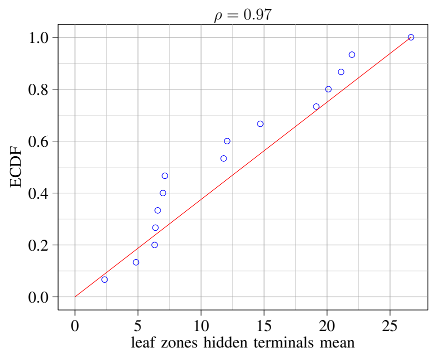


Figure 3. ECDF of the hidden-terminals' mean of the leaf zones.

and 13 show a pictorial view of the improvement of goodness of the power law fitting of the core (figure 13) versus base graphs (figure 12). These results validate the methodology proposed in [1] for the topology generation of rural areas consisting of removing the terminal nodes.

It is interesting that the rank exponent of the core-graphs is -0.64 with a coefficient of variation of only 15%. This fact suggests that a rank exponent around -0.64 is a characteristic of this type of graphs. Regarding the hop exponent, table I (b) shows that the power law approximation is very accurate for $h \ll \delta$ ($\rho \approx 1$ in all cases). Figure 6 shows the goodness of this fitting. However, table I (b) shows that it has $c_v = 23\%$, thus, having a higher variability than the rank exponent.

Regarding the hidden-terminals, table I (b) shows the goodness of the gamma distribution fitting (ρ is in the range $[0.93, 1]$). Figure 8 gives a pictorial view of this fact. The values of the shape and rate parameter of the gamma distribution vary significantly ($c_v = 0.63$ and $c_v = 1$, respectively). The value $c_v = 1$ for the rate parameter suggests it is exponentially distributed. This is confirmed by the semi-log plot of the CECDF of the rate parameter depicted in figure 2.

Figure 3 shows the ECDF of the hidden-terminals' mean of the leaf zones. The figure shows that the distribution is well fitted by a uniform distribution. Additionally, a Pearson's Chi-squared test of independence of the mean and the rate parameter of the leaf zones gives a p-value equal to 0.23. Thus, the independence hypothesis of these parameters cannot be discarded. This fact suggest the following method to generate

zone	μ_1	μ_2	θ	μ_T	ρ
1	0.80	5.10	0.87	1.37	1.00
2	0.49	1.86	0.64	0.99	1.00
3	1.09	4.35	0.67	2.18	0.99
4	0.33	2.19	0.67	0.95	0.99
5	0.43	2.39	0.70	1.02	1.00
6	0.68	2.75	0.67	1.38	0.99
7	0.30	7.83	0.54	3.73	0.99
8	0.63	0.63	1.00	0.63	0.99
9	2.35	2.35	1.00	2.35	0.99
10	0.79	3.20	0.72	1.48	0.99
11	0.73	2.23	0.28	1.81	0.99
12	0.82	4.12	0.66	1.95	1.00
13	0.51	2.25	0.65	1.11	0.99
14	0.23	1.41	0.21	1.16	0.99
15	0.33	1.03	0.53	0.66	0.99
16	2.38	2.38	1.00	2.38	1.00
μ	0.81	2.9	0.67	1.6	
c_i	0.35	0.94	0.12	0.43	
σ	0.65	1.8	0.23	0.81	
c_v	0.81	0.61	0.34	0.51	

Table II
LINK LENGTH DISTRIBUTION FITTING (μ_1, μ_2 AND μ_T ARE IN KM).

the hidden-terminals distribution of guifi.net zones:

1) Choose the mean number of hidden-terminals of the zone, $\langle r \rangle$, uniformly distributed between 0 and the desired maximum.
2) Choose the rate parameter of the gamma distribution of the hidden-terminals distribution of the zone, β, exponentially distributed with parameter $\lambda = 0.087$.
3) Choose the shape parameter of the gamma distribution of the hidden-terminals distribution of the zone as $\alpha = \langle r \rangle \beta$.

Comparing the figures of Catalunya zone with the leaf zones, we obtain the same conclusions as for the rural area zones: (1) Upon removing the terminals, the core-graph is well fitted by a power-law (figure 5). (2) The terminals of the base graph are gamma-distributed (figure 8 (a)). This fact shows that there is some self-similarity in the topology of guifi.net.

IV. Link Length Distribution

We have found that the link length distribution can be approximated by a mixture of 2 exponentials. Let L be the complementary CDF of the link length, X, then:

$$L(x|\lambda_1,\lambda_2,\theta) = P(X > x) = \theta\, \mathrm{e}^{-\lambda_1 x} + (1-\theta)\, \mathrm{e}^{-\lambda_2 x} \quad (1)$$

Table II summarizes the parameter fitting. In the table are given the link length means in km: $\mu_1 = 1/\lambda_1$, $\mu_2 = 1/\lambda_2$, $\mu_T = \theta\, \mu_1 + (1-\theta)\, \mu_2$. The entries have been organized such that $\mu_1 \leq \mu_2$. The last rows of table II give the sample mean (μ), its 95% confidence interval (c_i), the standard deviation (σ) and the coefficient of variation ($c_v = \sigma/\mu$) of the values corresponding to the leaf zones (all but 1: Catalunya) in the respective columns.

The sample correlation coefficient (ρ) in table II shows that the mixture of two exponentials gives a good fitting of the link length distribution. This can also be observed in figures 7 and 15, which show the link length distribution of Catalunya and leaf zones, respectively. An intuitive explanation of this result could be that links can be grouped in two sets: one set of short links characterizing connection of nodes located in closer geographical areas, for instance, villages in rural zones, or suburbs in Barcelona, and another set formed by longer distance links interconnecting nodes from different groups of short links. For instance, if we consider Catalunya zone (first row in table II), we have that 87% of nodes belong to the first group ($\theta = 0.87$), with a mean link length of $\mu_1 = 0.8$ km. The remaining 13% belong to long distance links with mean $\mu_1 = 5.1$ km.

Table II shows that the mean link lengths, and proportion of links that belong to the group of short and long distance links varies significantly from one zone to another. This is a logical result, since the groups of short and long distance links depend on may factors: population clouds, topography of the area (e.g. the existence of high antennas where many nodes can connect), etc.

V. Conclusions

This report is a companion of paper [1]. The objective was validating the conclusions obtained in [1] by considering a larger number of guifi.net zones. To do so, the same graphs obtained in [1] have been reproduced for 15 leaf zones, and Catalunya zone, which is formed by the aggregation of a number of leaf zones.

The numerical results shown in this report are completely in line with those obtained in [1]: None of the rural zones of guifi.net was well fitted by a power law. On the other hand, by removing the terminal nodes we obtain a *core-graph* which is in all cases reasonably well fitted by a power law. Interestingly, in all cases the number of terminal nodes is very well fitted by a gamma distribution. These results validate the topology generator proposed in [1] for this type of zones. Catalunya zone has been also found to satisfy these properties. This suggest some degree of self similarity on the topology of guifi.net.

Finally, we have investigated the link length distribution. Numerical results show that it is well fitted by a mixture of two exponentials.

Acknowledgments

This work was supported by the European project CONFINE http://confine-project.eu, and Spanish grants TIN2010-21378-C02-01 and 2009-SGR-1167.

References

[1] Llorenç Cerdà-Alabern. On the topology characterization of guifi.net. In *2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'2012)*, pages 389–396, Barcelona, Spain, October 8–10, 2012.

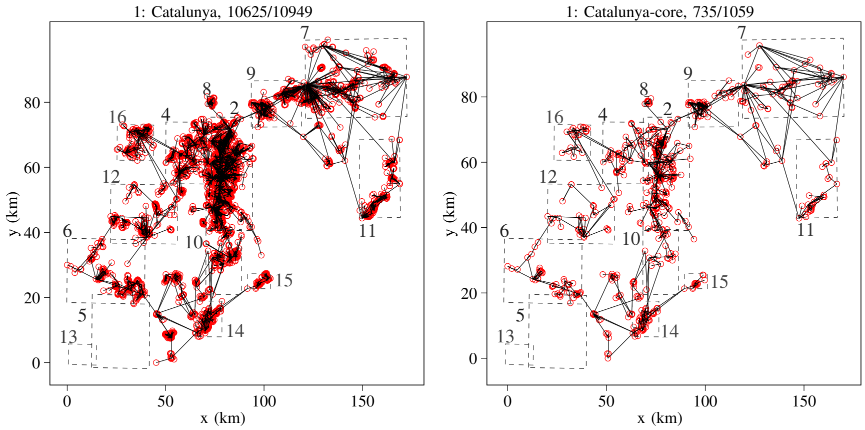


Figure 4. Base and core graphs of Catalunya zone. Axes are in km.

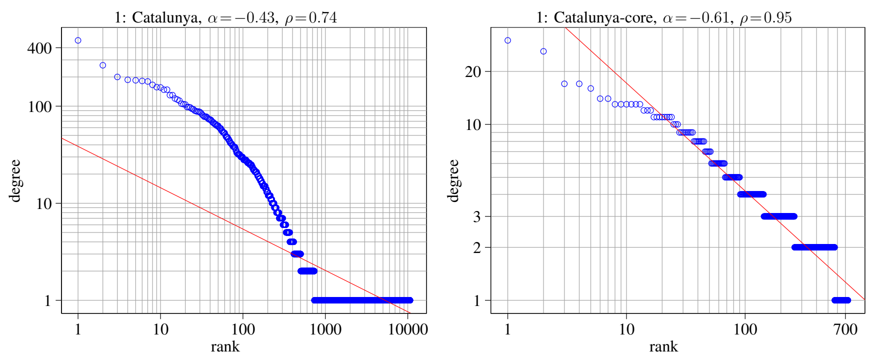


Figure 5. Rank $\log_{10}$-$\log_{10}$ plot of base and core graphs of Catalunya zone.

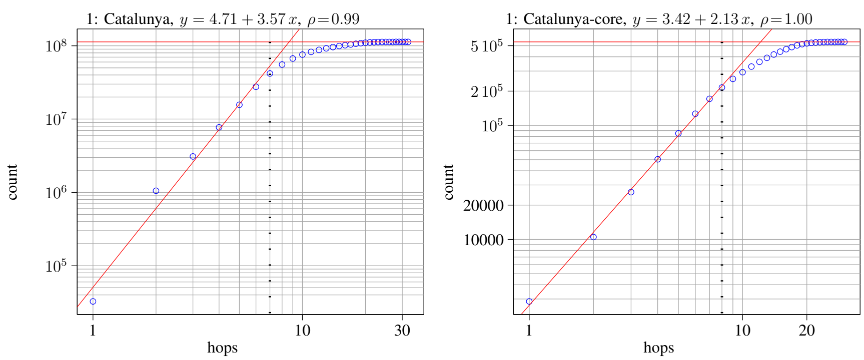


Figure 6. Hops count $\log_{10}$-$\log_{10}$ plots of base and core graphs of Catalunya zone.

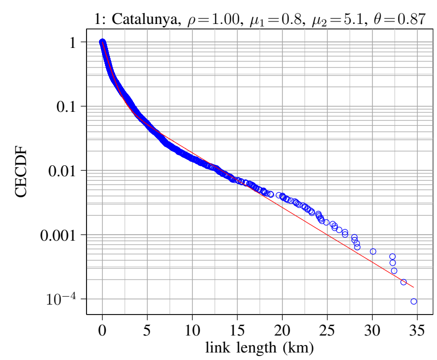
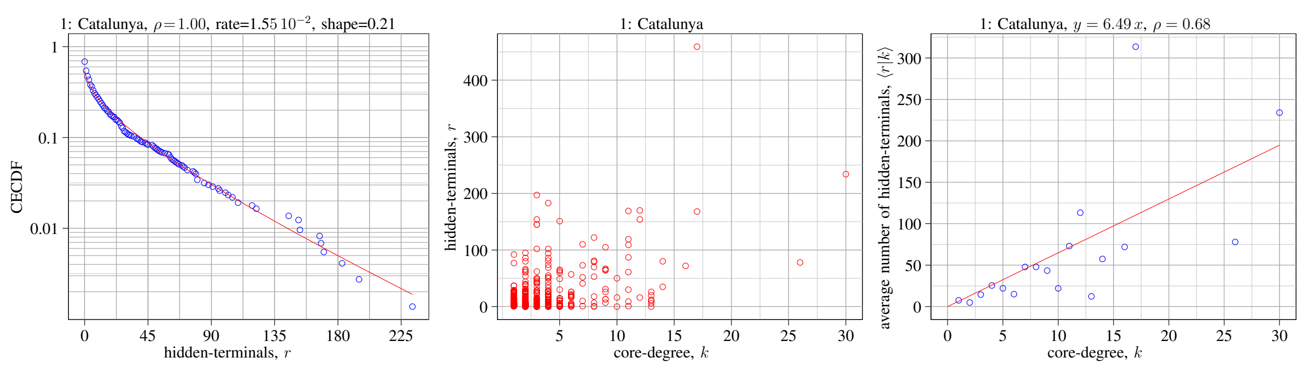


Figure 7. Link length distribution of Catalunya zone.



(a) (b) (c)

Figure 8. Catalunya zone core graph: (a) Hidden-terminals CECDF semi-$\log_{10}$ plot; (b) hidden-terminals vs core-degree scatter plot; (c) average number of hidden-terminals vs core-degree.

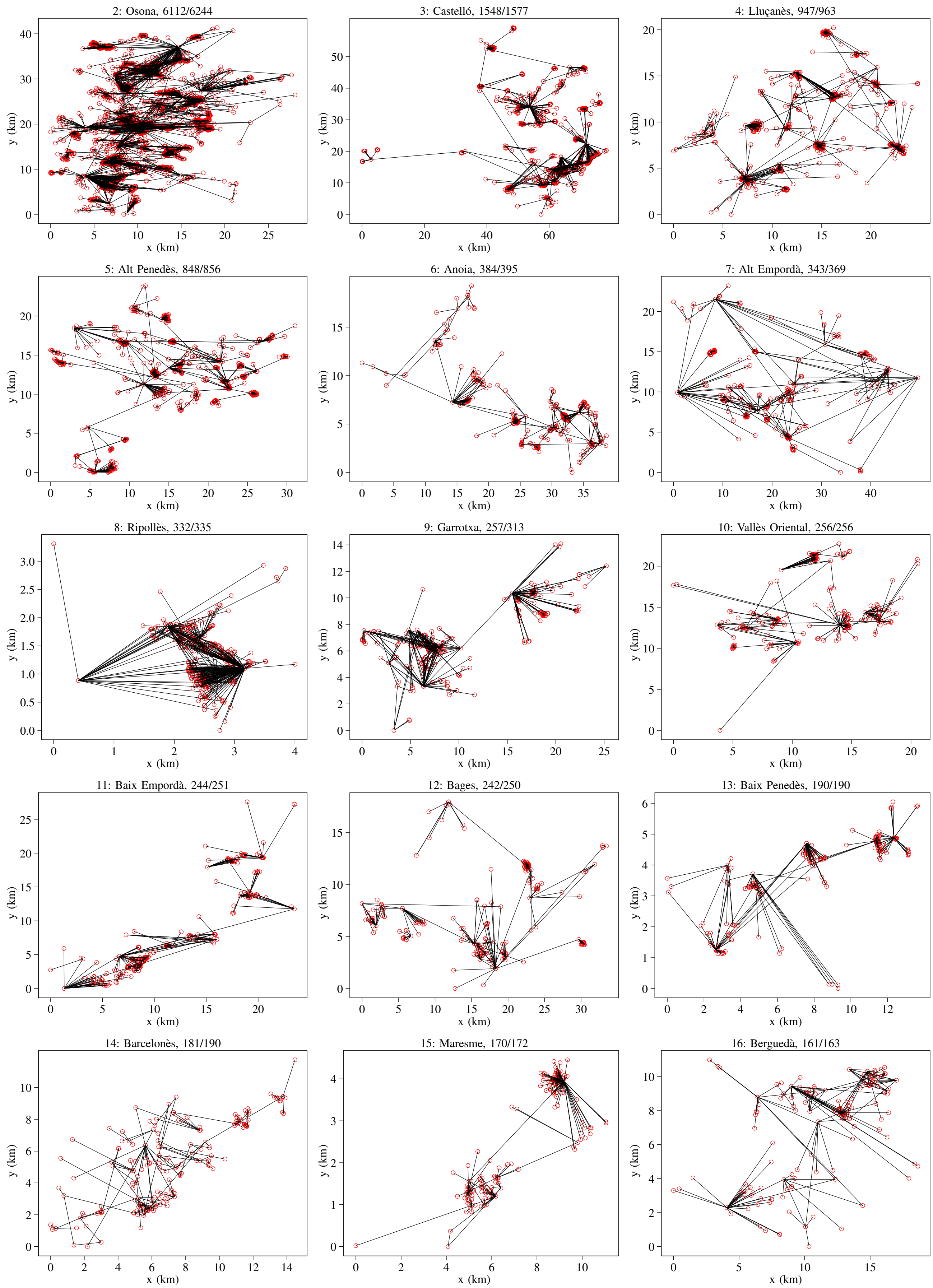


Figure 9. Base-graphs of the leaf zones. Axes are in km.

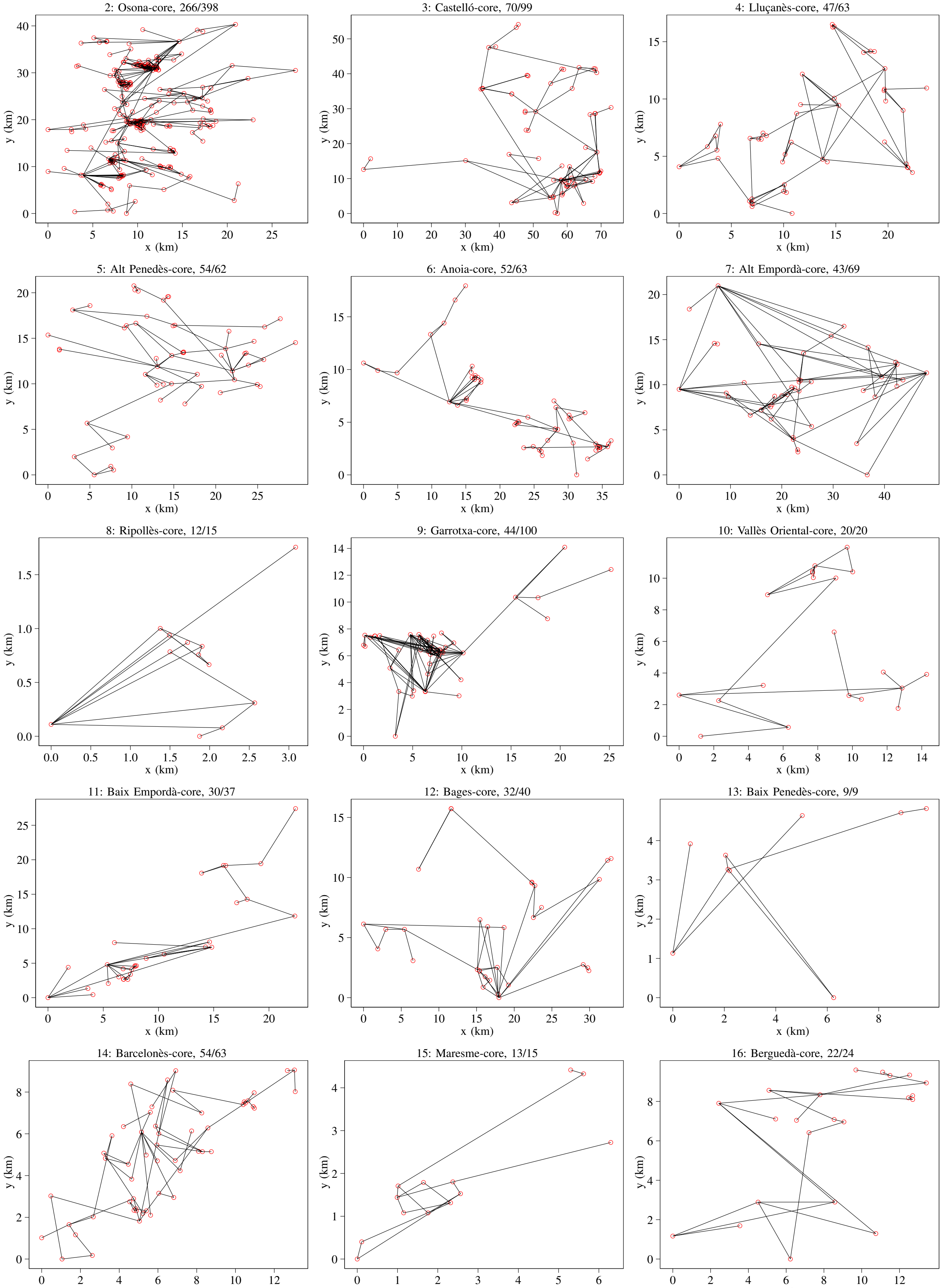


Figure 10. Core-graphs of the leaf zones. Axes are in km.

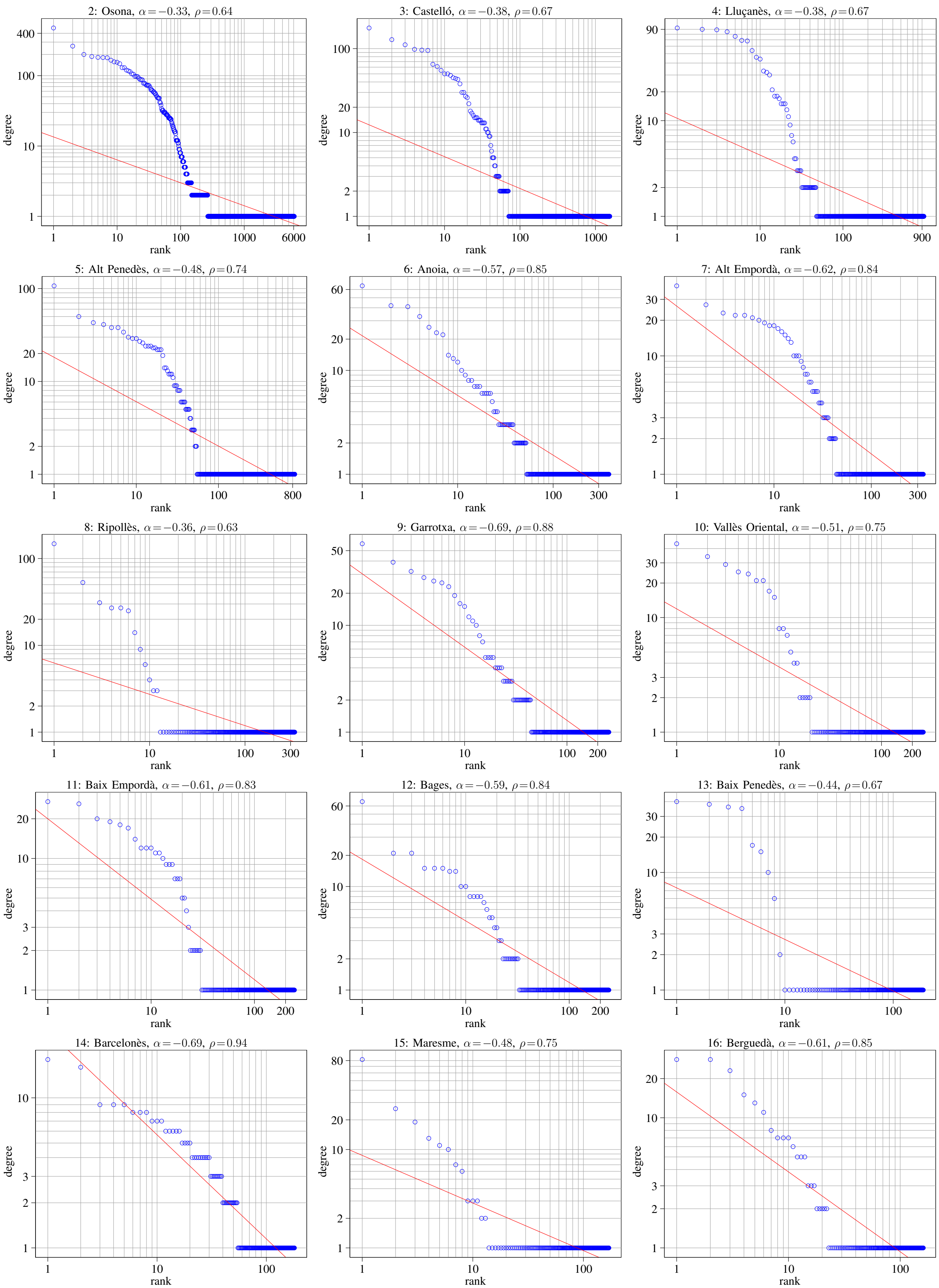


Figure 11. Rank $\log_{10}$-$\log_{10}$ plots of the leaf zones.

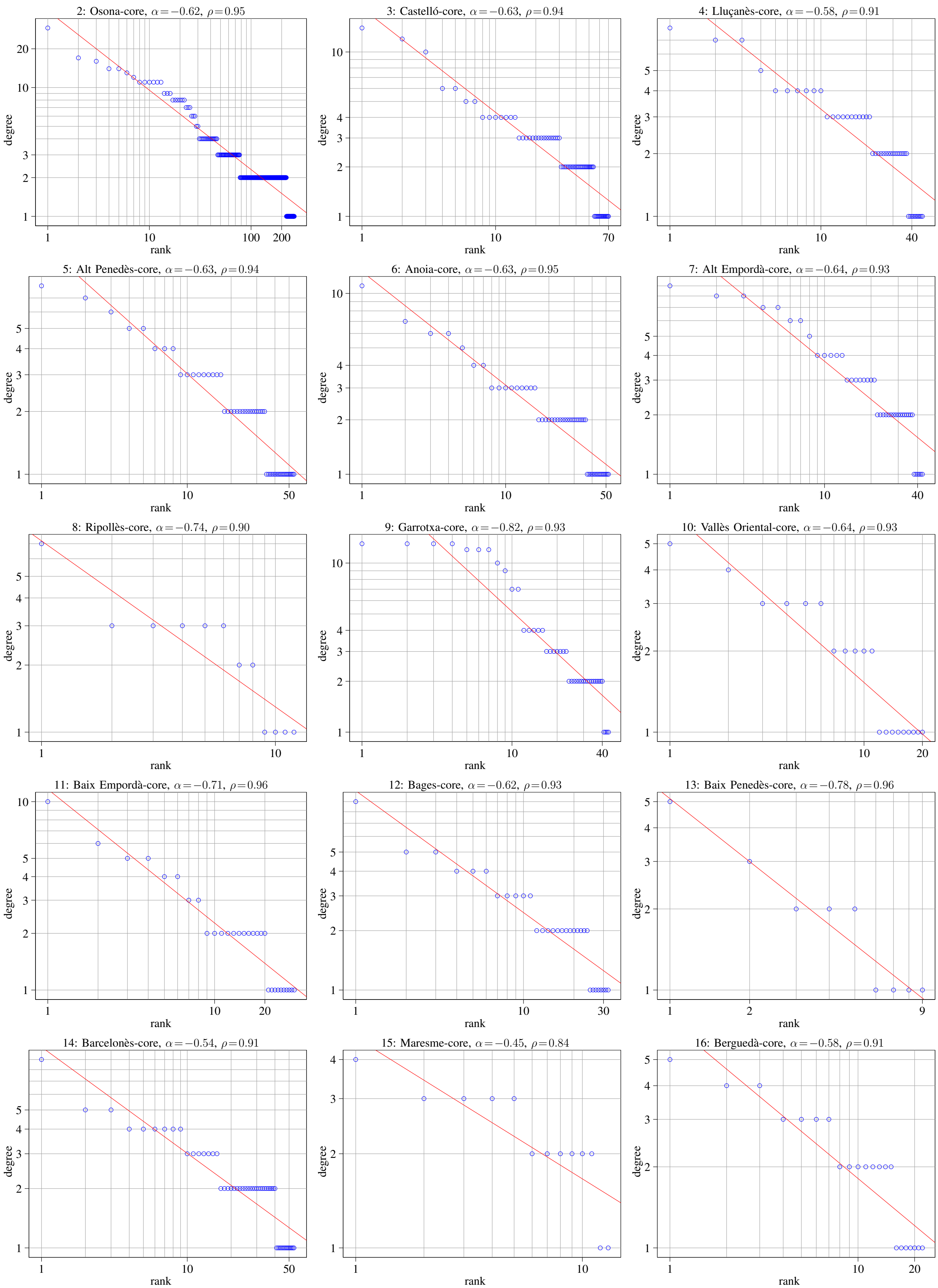


Figure 12. Rank $\log_{10}$-$\log_{10}$ plots of core-graphs of the leaf zones.

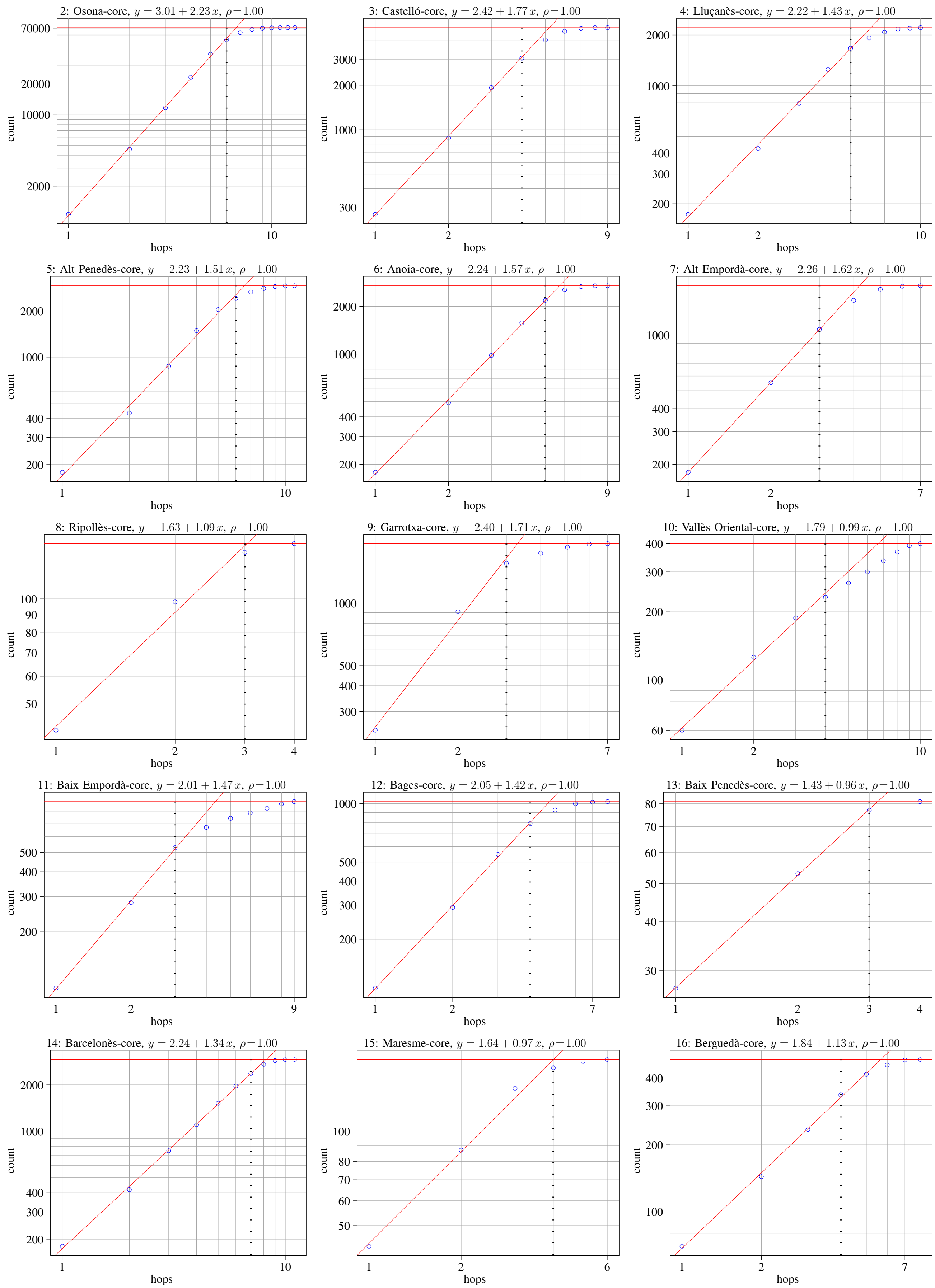


Figure 13. Hops count $\log_{10}$-$\log_{10}$ plots of core-graphs of the leaf zones.

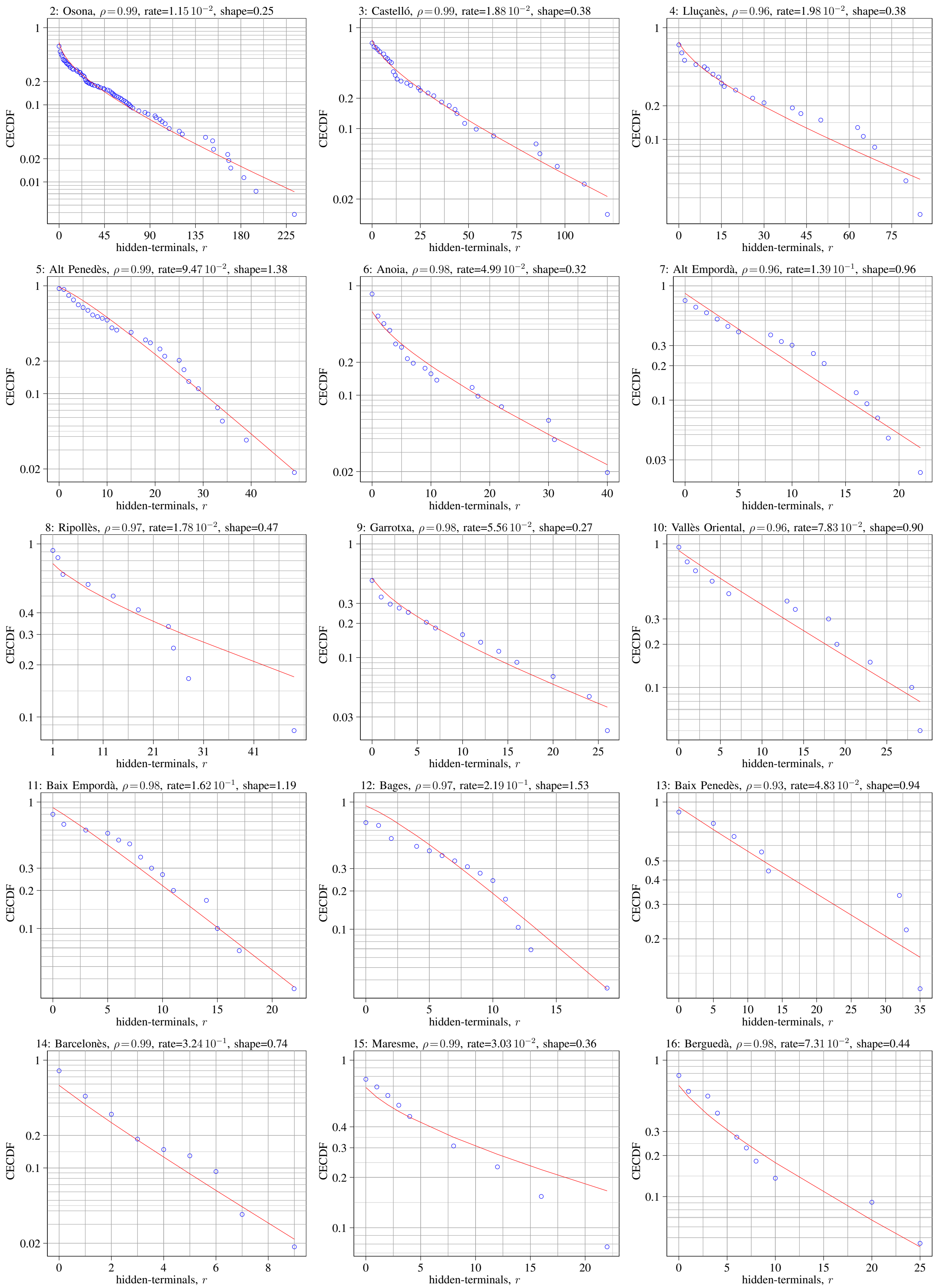


Figure 14. Hidden-terminals CECDF semi-$\log_{10}$ plots of the leaf zones.

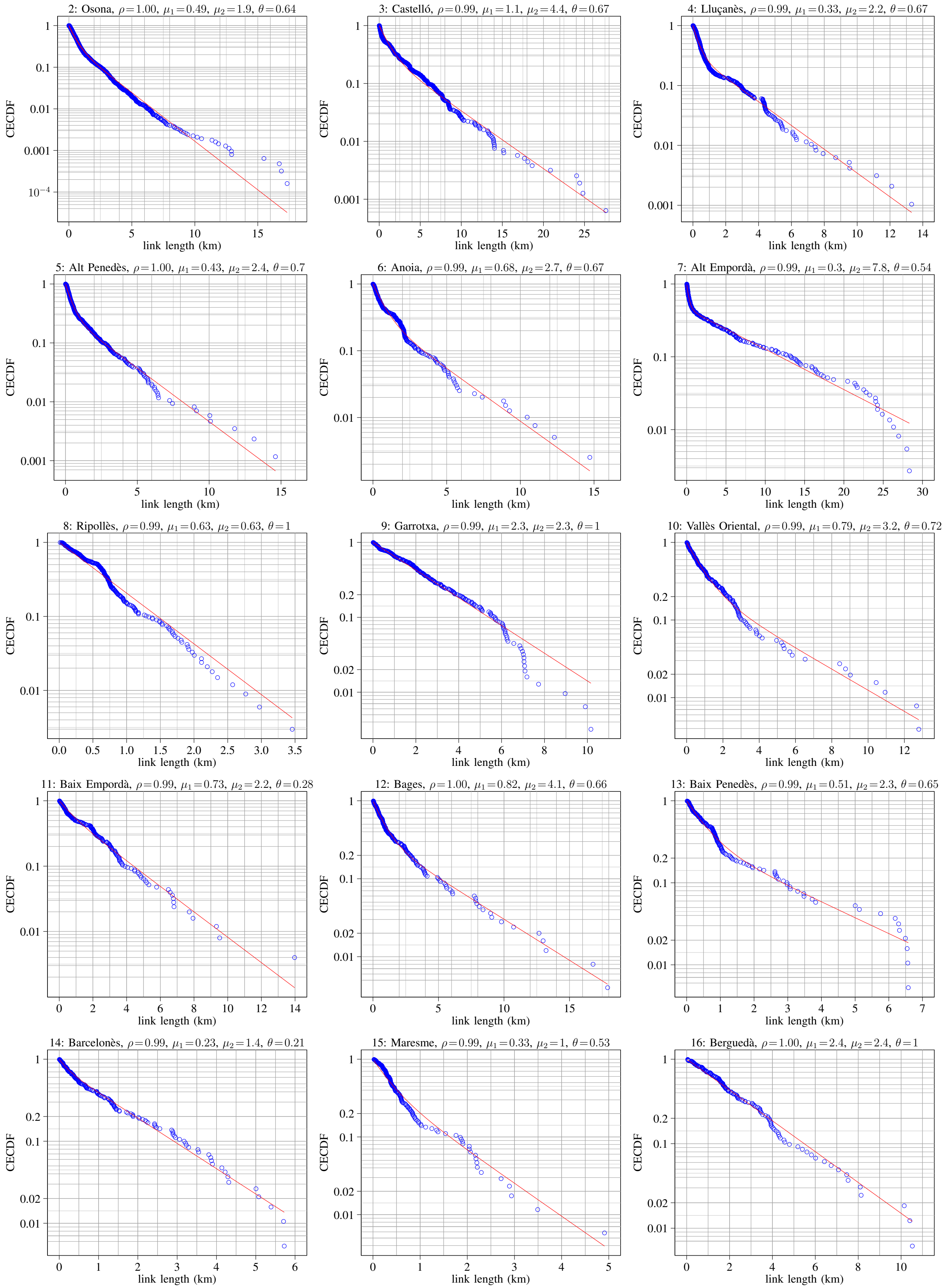


Figure 15. Link length CECDF plots of the leaf zones.