

On the Guifi.net Community Network Economics

Llorenç Cerdà-Alabern, Roger Baig and Leandro Navarro

July 27, 2019

Abstract

How costs are distributed among the participants is a key question in the management and viability of shared resources. Although all cost-sharing mechanisms are subjective and thus it is eventually up to the participants to accept one or another, some general criteria seem desirable, such as being budget-balanced and that, in any case, a participant pays more when not cooperating with anyone else. In this paper, we analyse the cost-sharing mechanism that the Guifi.net community network has developed and put in practice to split the transit costs among their more than 20 beneficiaries for almost a decade. To this end, we first introduce the general context and present the particularities of our case study, for which we elaborate a mathematical model including the characteristic cost function and cost-sharing mechanism. Then, we evaluate the model for a specific dataset, and we compare the results with the Shapley value, a reference cost-sharing mechanism in many areas. We follow with an investigation on the sensitivity of the system to reductions of the number of participants and demand. We conclude with some observations derived from our work for communities of practice to consider to face what seem to be recurrent periods of internal tension in which cooperation is endangered either because some participants cast doubt on the cost-sharing criteria or because they think that they would obtain a better payoff outside the coalition. The considered cost categories include network service providers, Internet exchange points, housing, electrical power, and human resources. Our results show that the Guifi.net's cost-sharing mechanism of the external connectivity, which comprises an equal membership fee for each participant plus a proportional distribution of the remaining costs according to the resource consumption, yields a cost assignment similar to the Shapley value. Our analysis also shows that any alternative to the coalition of all participants entails significant total cost increases and detrimental widespread cost allocation. The quality of the data that are made available to us has allowed us to reach an uncommon depth of analysis. We have made a great portion of this information publicly available to ease the validation of our results and to enable further investigation.

1 Introduction

Cost sharing allows the development and operation of a few common resources among multiple parties instead of building and maintaining multiple separate facilities. The potential for efficiency and sustainability in sharing comes with the overhead of coordinating the participation, contribution, and output allocation among multiple participants with widely different possible capacities and needs. Thus, efficient cost allocation methods are required.

In this paper, we address the challenge of advancing the development of such methods in the particular case of network services and infrastructure because the potential for savings due to cooperation in this field has been well described in [20]. Nonetheless, the lack of case studies seems to be a major barrier for the uptake of resources and service-sharing practices beyond Internet exchange points (IXPs) or some community networks. To contribute to filling this gap, we analyse in detail the mechanisms developed and put in practice by the community of practice of Guifi.net [2] to jointly enable the provisioning of a critical service, such as the transit service, and to share its costs.

Guifi.net, the largest community network in the world, started in 2004 as a grass-roots movement with the objective to provide Internet access to rural areas using WiFi. However, over time, Guifi.net experienced enormous growth, offering novel models of participation and new business opportunities. This environment has stimulated the development of Internet service providers (ISPs, *operators* in Guifi.net's terminology) that offer professional services inside the network and make investments to improve and extend the infrastructure and services. At the time of writing, the investment in optical fibre has clearly exceeded the investment in WiFi technology, and it is expected that the fibre rollouts in the middle term will have replaced almost all WiFi deployments.

The capacity to enable economic local and regional ecosystems with professionals offering services has proved to be an essential aspect for the success and growth of Guifi.net. In these ecosystems, the professionals compete with each other for the customers, while collaborating to build a common infrastructure that allows them to develop their professional activity. A key component for their proper development is the cost-sharing mechanisms used to distribute the costs of building and operating the shared resources among the beneficiaries who make significant use of them. These cost-sharing mechanisms are referred to as Guifi.net's cost-compensation system. The beneficiaries are participants because they are not limited to those who undertake professional activities through the shared infrastructures but also include associations of users, entities, and small and medium enterprises who provide themselves with connectivity and services. The development and implementation of these cost-sharing mechanisms have contributed to building trust and reducing conflict. This has had a very positive influence on increasing investment.

The Guifi.net cost-compensation system as a whole was described in [3]. Here, we focus on its implementation to distribute the costs of the interconnection service between Guifi.net and the rest of the Internet, which is the network service with the largest number of participants. We analyse this case in sufficient depth to show the traits and operational details of a working implementation of one specific cost-sharing mechanism. The example shows how equally distributing part of the costs acts as a protection against overuse by participants with irrelevant partaking and how proportionally splitting the rest by resource consumption (the exterior traffic in this case) contributes to lower entry barriers for low-demand participants without undermining the incentives for cooperating for those with the largest demands. The rest of the cost-sharing mechanisms of Guifi.net's compensation system are structured similarly.

The novel contributions of this paper are the following:

- A detailed description of the use case of the cost-compensation system developed by the Guifi.net community of practice.
- The elaboration of a mathematical model of the characteristic cost function and the cost-sharing mechanism.
- A comparison with the Shapley value, a well-known cost-sharing mechanism with which Guifi.net's mechanism has some interesting similarities.
- A sensitivity analysis regarding high-demand participants and variations of the cost-sharing mechanism parametrisation.

The close collaboration our research group has had with the Guifi.net community for almost a decade and the amount of information publicly available resulting from the model chosen and developed to deploy and operate the network infrastructures and services, in conjunction with detailed datasets provided to us by the Guifi.net Foundation [8] have enabled an uncommon depth of analysis.

The rest of the paper is structured as follows. In Section 2, we review the relevant related work. Section 3 gives a general overview of the network architecture of Guifi.net as a whole and introduces some technical details of the network operation centre (NOC) required for the next sections. In Sections 4 and 5, we discuss the general criteria for cost classification and Guifi.net's cost-sharing

fundamentals. In Section 6, we present the datasets we used and some of the initial findings. In Section 7, we elaborate on the various components of the cost model that we propose, and in Section 8, we formalise Guifi.net’s cost-sharing mechanism and compare its results with the Shapley value. In Section 9, we investigate the influence of desertions. In Section 10, we further investigate the relationship between the Shapley value and Guifi.net’s cost-sharing algorithm, and in Section 11, we make a proposal for improving Guifi.net’s method. Finally, in Section 12, we confront our findings with the current existing debate within Guifi.net and present our conclusions in Section 13.

2 Related work

Sharing or working together leads to the problem of how to divide the joint costs and cost savings among participants. As [22] stated, no single absolutely best set of criteria exist, but diverse policies are aimed to treat every participant in equivalent terms, being fair, impartial, and just without favouring or discriminating any party. The policy choice depends on a number of factors, including strategic aspects of cost allocation situations and participant preferences.

The 95-percentile method [6], which is billing based on the 95th percentile of traffic volumes sampled over five-minute intervals (also known as burstable billing) [24], is the most standard measure for billing traffic in ISPs and transit providers. It is an indicator of the network usage used for the dimensioning of the network infrastructure. The Shapley value [19] is an established way to fairly distribute gains and costs among several actors working in coalition. [21, 13] uses the Shapley value to compute the fair contribution of each flow to the 95th percentile price of interconnected links. As noted in [20], Internet transit is a highest cost for an ISP, and aggregating transit costs and group buying (*tuángòu*) decrease the cost due to the effect of economies of scale of typical sub-additive pricing and burstable billing because not all ISPs transit their peak traffic in the same period.

There are many experiences of cost sharing in IXPs [9] and community networks, such as B4RN [1], Guifi.net [2, 3], or RemIX [23], to achieve quality and affordability of backhaul connectivity in remote and underserved regions and therefore ensure the autonomy and sustainability of small ISPs. In [3], we examined some aspects related to the last-mile economics based on the monthly records of four WiFi and optical fibre deployments of the two previous years. However, beyond the case of Guifi.net, these cooperative operators have limited or incomplete traffic and economic log data, little formal and detailed cost analyses, and therefore, a lack of cost fairness optimisation. Beyond this, commercial network providers consider the detailed traffic and pricing data to be commercially sensitive and confidential and, therefore, inaccessible for analysis.

3 Network architecture

Figure 1 shows the Guifi.net general network architecture. Currently, almost all the traffic between Guifi.net and the rest of the Internet (the external traffic) is exchanged at the premises of Equinix [7], the main collocation centre in Barcelona, where Guifi.net has its NOC. The traffic is exchanged preferably through peering in the local IXP, the CATNIX [4] in our case, and otherwise through network service providers (NSPs), also referred to as *carriers*. At the time of this writing, Guifi.net has five NSPs, two for general purpose Internet connectivity, which account for over 99% of the NSP traffic, and three for special operations, such as specific peering or operations related to mobile and fixed telephony services. The setting of two general purpose Internet connectivity providers ensures basic redundancy. We refer to these carriers as C_1 and C_2 . To ensure a certain quality of service in the case of failure, the hired capacities of C_1 and C_2 are required to be always the same.

The pricing scheme of C_1 is a strict flat rate for its entire capacity, and the pricing for C_2 is a flat rate for the traffic below 10% of its capacity plus a per traffic fee for the traffic exceeding this. The flat rate part of C_2 is around two-thirds the cost of C_1 . The routing policy is to maximise the use

of the capacity of C_1 within a safety margin, and the rest of the traffic is diverted to C_2 . The hired capacities are extended when the cost of the per traffic fee of C_2 becomes greater than the cost of the capacity expansion of C_1 . On average, the contracts are being renegotiated every year or year and a half to accommodate the growth in demand. In May 2019, C_1 and C_2 have been upgraded from 2×10 Gbps to 3×10 Gbps, and the IXP connection is 1×10 Gbps, but it will be soon upgraded to 2×10 Gbps. The costs of the additional NSPs for specific operations is less than 4% of the total external connectivity costs.

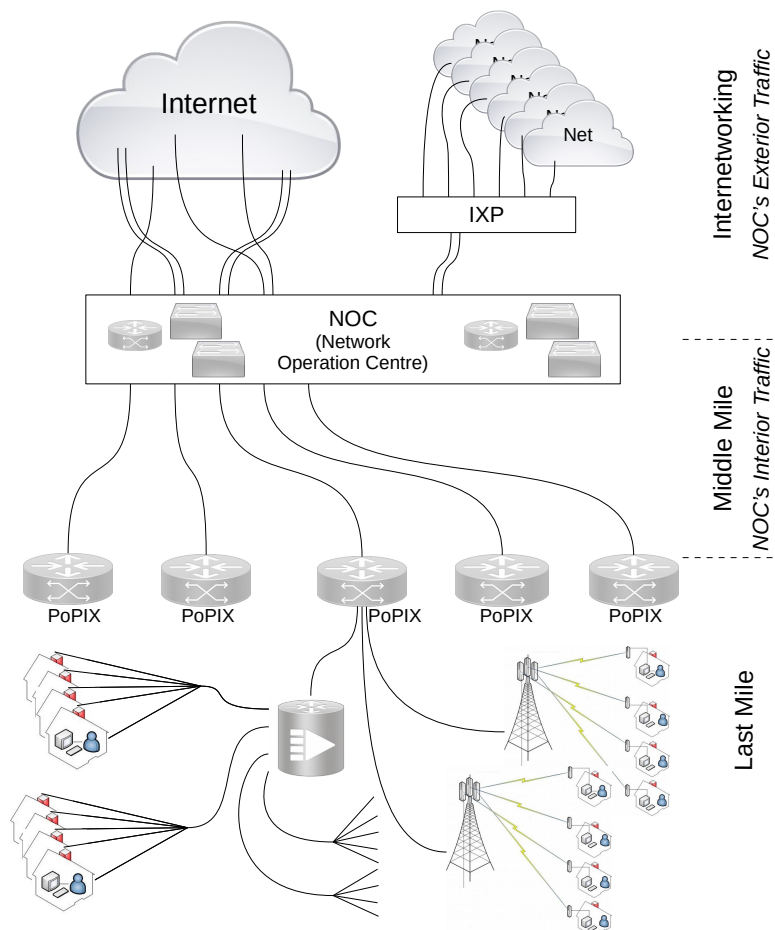


Figure 1: The Guifi.net network architecture.

The external traffic is distributed through the middle-mile network to the local points of presence (PoPIX) and from there to the end users through the last mile. The middle-mile network is basically built on third-party fibres (usually lit fibres). The last mile comprises a combination of owned WiFi and fibre optic, with the last replacing the first in existing deployments and is dominant in new deployments. Around 30 participants contribute to sharing the costs of the exterior traffic service and the middle infrastructure.

4 Cost classification

In Guifi.net, any cost to be compensated is split into one or more *clearing houses*. A clearinghouse is a functional aggregation of network components or services. At present, there are two categories of clearing houses in Guifi.net: the NOC, which includes all the services and infrastructure from the exterior traffic to the links connecting the NOC to the PoPIX, and the territorial, which encompass the last-mile services and infrastructure, that is, the infrastructure and its related services from the PoPIX to the final points of consumption. The costs of a cabinet (hardware, electricity bills, etc.) to

Item	Cost center			MAP [months]
	Exterior t.	Interior t.	Rack s.	
Hardware (routers, switches, etc.)	✓	✓	–	36
Insurance (annual fees)	✓	✓	✓	12
Administration	✓	✓	✓	–
Management	✓	✓	–	–
Technical support	✓	✓	–	–
Legal support (outsourced – monthly fee)	✓	✓	✓	–
Housing – acquisition cost	–	–	✓	36
Housing – monthly fee (including electricity)	–	–	✓	–
Housing – shared hardware slots	✓	✓	–	–
IXP – monthly fee	✓	–	–	–
carriers – monthly fee	✓	–	–	–
Middle-mile circuit – acquisition cost	–	–	✓	36
Middle-mile circuit – interconnection costs	–	–	✓	36
Middle-mile circuit – monthly fee	–	–	✓	–

Table 1: Most significant network operation centre costs, cost centre assignment (✓: assignment applies, –: assignment does not apply) and maximum amortisation period.

distribute the fibre in a district is an example of a cost to be allocated to a single clearinghouse, the corresponding territorial clearing house in this case. In contrast, some of the human resource costs of the Guifi.net Foundation are distributed among several clearinghouses because the workers perform tasks in them.

In the clearing houses, the costs are grouped by *cost centres*. For the NOC, the current cost centres are *exterior traffic*, *interior traffic*, and *rack spaces*. The exterior traffic costs include the NSPs and IXP, and the interior traffic includes the middle-mile links. Rack space costs are those that refer to the housing facilities. Table 1 shows the NOC’s most significant costs, how these are distributed among cost centres, and the maximum amortisation period (MAP) in months, if any.

Given that, in the NOC, the operational expenditures (OPEX) are much higher than the capital expenditures (CAPEX), CAPEX costs are subsumed as periodic charges for a given number of months. In contrast, in the territorial clearinghouses, CAPEX and OPEX cost compensations are calculated separately, not only because CAPEX is significantly higher but also because CAPEX calculations are significantly more complex because loans and other debts to be repaid are involved.

5 Cost-sharing fundamentals

The total cost borne by a participant of a clearinghouse is the result of a fixed charge equal for each participant, the *membership fee*, plus a charge for consumption for each cost centre from which the participant uses resources, the *consumption fees*. Consumption fees are usually calculated by proportionally allocating the costs of the cost centres among the resource usage of the participants.

The rationale behind this cost-sharing mechanism is two-fold. On one hand, the membership fees, which are equivalent to equally weighting part of the costs, act as a filter against opportunistic participants to keep the costs related to the number of participants under control by establishing an entry barrier. On the other hand, proportionally assigning most of the costs to each participant’s usage not only avoids negative externalisations from big consumers towards the smaller consumers but also has a wealth redistributive role because it makes technologies available that smaller consumers could afford neither separately nor through coalitions. Through this paper, we refer to a *coalition* for any subset of participants.

Two resources can be consumed at the NOC: network bandwidth and rack space. The metric used to measure the usage of network assets is the sum of the 95th percentile of five-minute time averages of the inbound and outbound traffic of the given link per participant. Rack space is measured through the standard *rack unit*, which is half the minimum size. To ensure liquidity, the cost-sharing calculations, the *compensations* in Guifi.net’s terms, are done periodically. In the case of the NOC, they are done on a monthly basis.

6 Datasets

The Guifi.net Foundation has given us access to a traffic dataset and to some of the spreadsheets generated on a monthly basis to calculate the cost sharing and has provided some additional information on carrier offerings. To maintain confidentiality, the names of the participants and carriers have been kept anonymous.

6.1 Traffic

The traffic dataset contains i) the inbound and outbound aggregated traffic of each participant (i.e. the traffic exchanged by each participant in both directions with the IXP and the carriers and the other participants all together), and ii) the inbound and outbound traffic of each carrier and the IXP. The dataset we have evaluated in this paper consists of the mean traffic of five-minute samples, recorded during the second week of April 2019, from Saturday the 6th to Friday the 12th. The data files are in CSV format and can be found in the public repository [11]. As an example of this data, Figure 2 shows the *inbound* and *outbound* traffic of three participants. The dashed lines correspond to their 95th percentiles. Throughout the paper, we sorted the participants in ascending order by the addition of the *inbound* and *outbound* 95th percentiles. Accordingly, the graph on the top corresponds to the participant with the least usage (26 Mbps), the one in the middle to medium usage (49 Mbps), and the one on the bottom to the highest usage (7.9 Gbps). As Figure 2 shows, the traffic patterns become more evident as the traffic of the participants increase.

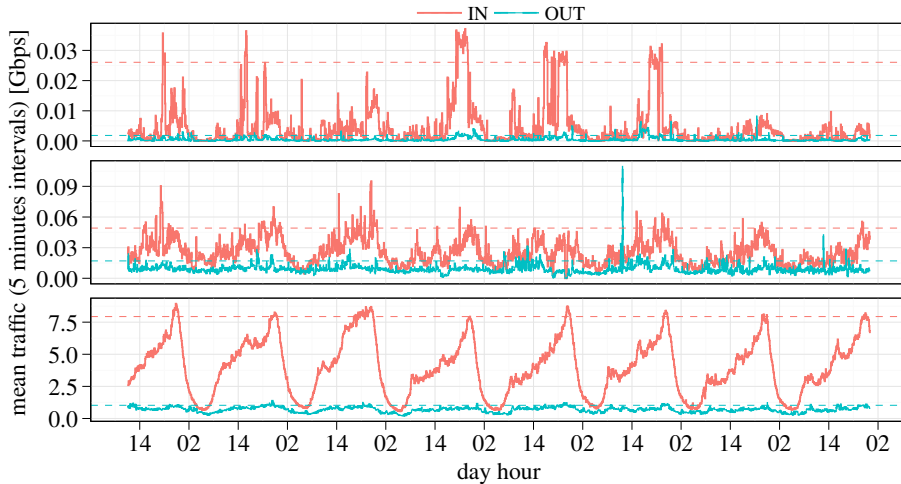


Figure 2: Exterior traffic of three participants. The dashed lines correspond to the 95th percentiles.

During the recorded period, 26 participants had traffic activity in the NOC. Figure 3 shows their mean traffic and 95th percentiles in the log scale sorted according to the aforementioned criterion. It is interesting to note that nearly a constant distance exists between the mean and 95th percentiles in the log scale, especially for participants with higher traffic. For instance, let q_p^{in} and m_p^{in} be the *in* traffic at the 95th percentile and the mean, respectively, of participant p . There are 13 participants using $m_p^{in} > 0.1$ Gbps. For these participants, the mean of q_p^{in}/m_p^{in} is 1.93 with a standard deviation

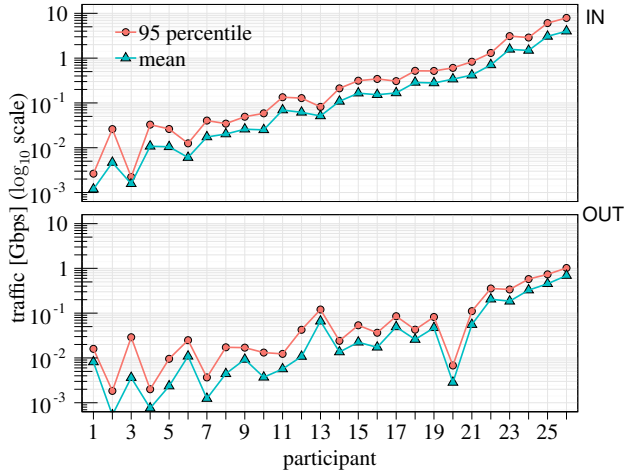


Figure 3: In and out traffic: mean and 95th percentile.

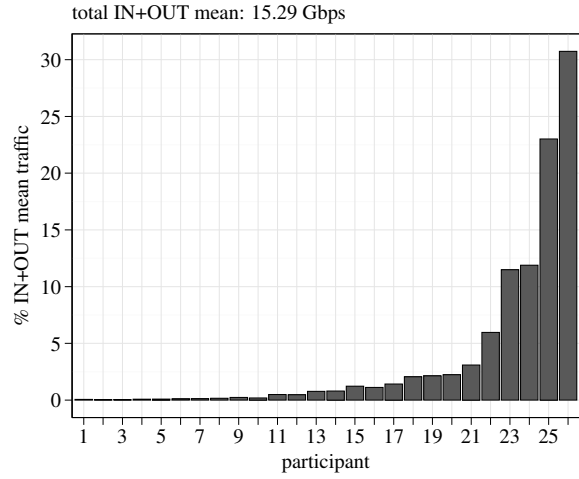


Figure 4: Percentage of mean of inbound and outbound traffic.

of only 0.11. Thus, if only the mean traffic is known, a rule of thumb for the five-minute sampling at the 95th percentile is taking twice the mean.

Outbound traffic that is higher than the inbound traffic corresponds to content providers, whereas the opposite indicates that the participant is either an end user or an access network provider. To appreciate the large difference between participants in terms of traffic exchanged, Figure 4 shows the percentage of mean inbound and outbound traffic exchanged by each participant. The participant with the highest usage exchanges 30.74% of all traffic, and the four participants with the highest usage account for the 77.12%. Indeed, all of them are ISPs, and together, they have over 25,000 subscribers. Conversely, the 10 participants with the least usage together exchange only 1.10% of the traffic. Among these participants, there are associations, self-provisioning participants, micro-ISPs, etc., that account for less than a thousand end users each.

6.2 Costs

The NOC’s cost compensation spreadsheet has numerous data and formulae because it is used to calculate the dues of all participants for the exterior traffic and the middle mile and racks. As already explained, for simplicity, we focus on the distribution of the costs associated with external traffic. To this end, we recalculated the cost assignments because, for historical reasons, in Guifi.net’s spreadsheets, the labour and other costs are entirely charged to the exterior traffic cost centre, whereas, as shown in Table 1, they should also be partially incurred by the internal traffic and, in some cases, by the rack spaces as well. To address this anomaly, we proportionally distributed the costs that should be incurred by more than one cost centre to the total amount of exclusive costs of each cost centre of the previous month. We used the data of the previous iteration because some of the exclusive costs, such as the cost per slot of housing, depend on the proportionality factors. Hence, they cannot be obtained from the data of the current iteration due to indeterminate problems.

	Cost centre			Total [€]
	Exterior t. [€]	Interior t. [€]	Rack s. [€]	
Exclusive	6,362.00	14,501.00	2,558.00	23,421.00
Shared	2,963.00	6,754.00	339.00	10,056.00
Total	9,325.00	21,255.00	2,897.00	33,477.00

Table 2: Considered costs per cost centre (April 2019).

Item	Cost [€]	Cost function category	
		NSPs & IXP	Other
C_1 (2×10 Gbps)	2,800.00	✓	–
C_2 flat rate (2×10 Gbps)	1,500.00	✓	–
C_2 excess	317.00	–	–
Additional NSPs	356.00	–	<i>equally split</i>
IXP (1×10 Gbps)	650.00	✓	–
RIPE-NCC	87.00	✓	<i>constant</i>
Cabling	652.00	–	<i>participants</i>
total	6,263.00		

Table 3: Exclusive exterior traffic cost breakdown (April 2019) and assignment to cost function categories (✓: assignment applies, –: assignment does not apply).

Table 2 shows the exclusive costs per each cost centre and our redistribution of the costs to be incurred by more than one cost centre and the corresponding totals. Table 3 shows the costs of the external traffic cost centre. Given that the deals with the carriers are private and subject to non-disclosure agreements, in this table and hereafter, we use approximations for the actual costs by rounding the values with deviations below 1%. The membership fee is 154.00€.

7 Cost estimation model

In this section, we derive a function to estimate the total cost of the external traffic on Guifi.net for an arbitrary coalition (i.e. a subset of participants) and traffic consumption. This will be the baseline cost function that we use in the cost-sharing analysis we perform later in Section 8. For ease of reference, Table 4 summarises the parameters and values defined in our model cost, and Table 5 displays the main variables of the paper.

Parameter	Value	Unit	Description
P	26	part.	Number of participants
A_P	$\{1, 2, \dots, P\}$	part.	Grand coalition
α_{C_1}	0.58	–	Proportion of traffic exchanged at C_1
α_{C_2}	0.17	–	Proportion of traffic exchanged at C_2
α_{IXP}	0.25	–	Proportion of traffic exchanged at IXP
K_{C_1}	1.25	–	C_1 overprovisioning factor
K_{C_2}	1.25	–	C_2 overprovisioning factor
K_{IXP}	1.5	–	IXP overprovisioning factor
$q(A_P)$	25.06	Gbps	95th-perc. of traffic exchanged by A_P , Eq. (3)
$m(A_P)$	15.29	Gbps	Mean traffic exchanged by A_P , Eq. (7)
c_T	8,575.00	€	Total cost of A_P
H_f	1,193.33	€	Other costs parameter, Eq. (6)
H_d	$\frac{1,423.33}{m(A_P)}$	€/Gbps	Other cost parameters, Eq. (6)
H_p	$\frac{1,758.33}{P}$	€/part.	Other cost parameters, Eq. (6)

Table 4: Parameters of the cost model.

In line with the classification in Table 3, we propose a cost function with components that can be grouped into two main categories: the *NSP and IXP costs*, which exclusively depend on the port

Variable	Description
i	External connectivity provider, $i \in \{C_1, C_2, \text{IXP}\}$
p	Participant, $p \in \{1, 2, \dots, P\}$
A	Coalition of participants, $A \subseteq \{1, 2, \dots, P\}$
q_i^{in}, q_i^{out}	95th percentile of <i>in</i> and <i>out</i> traffic of provider i
q_p^{in}, q_p^{out}	95th percentile of <i>in</i> and <i>out</i> traffic of participant p
m_p^{in}, m_p^{out}	Mean of <i>in</i> and <i>out</i> traffic of participant p
$q_{in}(A), q_{out}(A)$	95th percentile of <i>in</i> and <i>out</i> aggregated external traffic exchanged by A
$m_{in}(A), m_{out}(A)$	Mean of <i>in</i> and <i>out</i> aggregated external traffic exchanged by A
g_p	Guifi.net cost assigned to participant p
s_p	Shapley value cost assigned to participant p

Table 5: Main variables used in this paper.

capacities contracted to the NSPs and the IXP, and the *other costs*, such as housing, electricity, exceeding traffic, insurance, manpower, amortisations, etc., that is, costs that depend on several factors, with the number of participants and traffic volume being the most significant. Roughly, each of them represents one-half of the total cost. These categories must not be confused with the exclusive and shared cost classification criterion from Section 6.

7.1 NSP and IXP costs

The total cost of this category is the sum of the cheapest combinations of links offered by the NSPs and the IXP that are able to satisfy the respective capacity demands. Since the links are symmetric, each provider demand is established based on the maximum of its inbound and outbound traffic of the previous iteration, that is, of the traffic consumption of the previous month. For dimensioning purposes, the traffic to be considered is given by the following:

$$q_i = \max(q_i^{in}, q_i^{out}), \quad i \in \{C_1, C_2, \text{IXP}\}, \quad (1)$$

where q_i^{in} and q_i^{out} are respectively the 95th percentile of the *in* and *out* traffic of provider i .

Using Equation (1), we can compute q_i of the coalition containing all participants, the grand coalition of all participants. However, for smaller groups of participants (coalitions), we assume that the NSP and IXP traffic proportions remain the same because the dataset only encompasses aggregate traffic records but not which proportion of the traffic of each participant is exchanged at NSPs and the IXP. That is, we let P be the number of participants, and $A \subseteq \{1, 2, \dots, P\}$ be any coalition and let α_i , $i \in \{C_1, C_2, \text{IXP}\}$, be the proportion of traffic exchanged by all participants at C_1 , C_2 , and the IXP, respectively. Then, we obtain the following:

$$\alpha_i = \frac{q_i}{\sum q_i}, \quad i \in \{C_1, C_2, \text{IXP}\}, \quad (2)$$

and given aggregated external traffic exchanged by coalition A :

$$q(A) = \max(q_{in}(A), q_{out}(A)), \quad (3)$$

where $q_{in}(A)$ and $q_{out}(A)$ are respectively the 95th percentile of *in* and *out* aggregated external traffic exchanged by A , we assume that the traffic exchanged by A at NSPs and the IXP is $\alpha_i q(A)$, $i \in \{C_1, C_2, \text{IXP}\}$.

Establishing an overprovisioning factor K_i for each provider, we conclude that, for dimensioning purposes, the demands are the following:

$$d_i(A) = K_i \alpha_i q(A), \quad i \in \{C_1, C_2, \text{IXP}\}. \quad (4)$$

Table 6 shows the prices we considered. As already explained, NSP prices are subject to private negotiations and non-disclosure agreements. Thus, our precise information on the offers is limited to the current services contracted. The rest of the NSP values are estimations, as indicated. In contrast, the IXP fees are publicly available on its website [5].

Capacity	Monthly fees [€]		
	C_1	C_2	IXP
1 Gbps	*430.77	*276.92	200.00
10 Gbps	*1,400.00	*750.00	650.00
20 Gbps	2,800.00	1500.00	–
30 Gbps	4,200.00	1800.00	–
100 Gbps	*20,000.00	–	3,900.00

Table 6: Monthly fees of NSPs and the IXP (May 2019; * estimated values; – indicates no data available for that configuration and estimation is not needed).

Assuming cost additivity in the case of capacity aggregations, the minimum cost to allocate the traffic demand for each $i \in \{C_1, C_2, \text{IXP}\}$ is obtained by the solution of the integer linear program (ILP):

$$\begin{aligned}
 c_i(A) = \min \quad & \sum a_{ij} n_{ij}, \\
 \text{s. t.} \quad & \sum b_{ij} n_{ij} \geq d_i(A) \\
 & n_{ij} \geq 0, n_{ij} \in \mathbb{Z},
 \end{aligned} \tag{5}$$

where n_{ij} is the number of links of type j , with cost a_{ij} for a bandwidth of b_{ij} Gbps (rows in Table 6), and $d_i(A)$ is the demand given by Equation (4).

7.2 Other costs

We model these other costs as a three-term function. The first term gathers the fixed costs; thus, we model it as a constant. The second relates to the costs associated with the exchanged traffic, in this case, taking the addition of the means of inbound and outbound traffic, as shown in Equation (7). The third reflects the costs primarily influenced by the number of participants.

We estimate the other costs as follows:

$$c_h(A) = H_f + H_d m(A) + H_p |A|, \tag{6}$$

where $|A|$ is the number of participants in coalition A (i.e. the cardinality of A) and $m(A)$ is given by:

$$m(A) = m_{in}(A) + m_{out}(A), \tag{7}$$

where $m_{in}(A)$ and $m_{out}(A)$ are respectively the mean of *in* and *out* aggregated external traffic exchanged by A .

To assess the constants in Equation (6), we used the values in Table 2 and Table 3. We split the sum of the shared costs and the additional NSPs into three terms, and we added the RIPE-NCC fee to the constant, the C_2 excess to the traffic, and the cabling to the participant terms, respectively. This results in the values reported in Table 4.

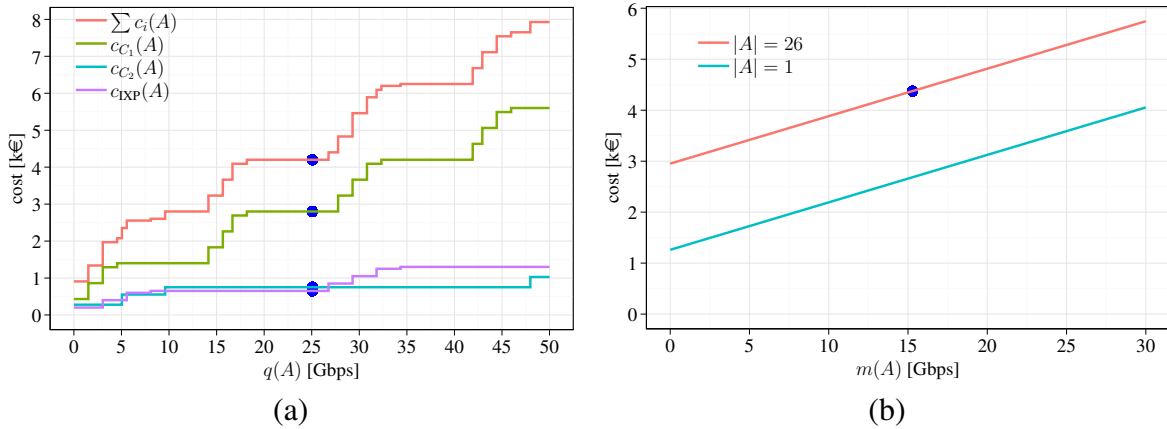


Figure 5: (a) NSP and IXP costs (Eq. (5)), and (b) other costs (Eq. (6)). The dots show the costs for the grand coalition.

7.3 Total cost of a coalition

Summarising, we estimate the overall monthly cost of the external traffic exchanged by any coalition $A \subseteq \{1, 2, \dots, P\}$ using the following equation:

$$c(A) = \sum_{i \in \{C_1, C_2, \text{IXP}\}} c_i(A) + c_h(A), \quad (8)$$

where $c_i(A)$ are the NSP and IXP costs given by Equation (5), and $c_h(A)$ includes the other cost given by Equation (6).

Figure 5 shows the NSP and IXP costs (a) and the other costs (b) using the model given by Equations (5) and (6), respectively. In Figure 5 (b), we show the resulting cost for a single participant, $|A| = 1$, and the cost for the total number of participants present in the current dataset, $|A| = 26$. The NSP and IXP costs account for 4,200.00€, the other costs for 4,325.00€ and the total for 8,575.00€. The difference between the calculated total and the total in Table 2 (9,325.00€) is because, for the calculation, we did not impose d_{C_2} to be equal to or greater than d_{C_1} as done in Guifi.net, as explained in Section 3. If $d_{C_2} = d_{C_1}$ were set to compute $c_{C_2}(A)$, Equation (8) would also give 9,325.00€.

8 Cost sharing

In this section, we formalise the cost-sharing mechanism used by Guifi.net, and we compare it with the Shapley value. Let P be the total number of participants, $P = 26$, and A_P the set of all participants. Using Equation (8), the total cost assigned to the set of all participants is as follows:

$$c_T = c(A_P). \quad (9)$$

8.1 Guifi.net's cost-sharing mechanism

As explained in Section 5, a portion of the costs are equally weighted among the participants, and the rest are allocated according to their resource consumption. We refer to these parts as the *fixed*, F , and *shared*, S , costs:

$$\begin{aligned} F &= \gamma c_T && \text{fixed cost} \\ S &= (1 - \gamma) c_T && \text{shared cost,} \end{aligned} \quad (10)$$

where $0 \leq \gamma \leq 1$. The fixed cost, F , is equally divided among all participants. For simplicity, the shared cost, S , is proportionally divided according to each participant's usage. Thus, the cost g_p

assigned to a participant p is given by the convex combination:

$$g_p = \frac{F}{P} + S \frac{u_p}{\sum u_p} = \frac{\gamma c_T}{P} + (1 - \gamma) c_T \frac{u_p}{\sum u_p}, \quad p \in \{1, 2, \dots, P\}, \quad (11)$$

where u_p is the bandwidth usage of each participant, defined as the addition of the 95th percentiles of its inbound and outbound traffic:

$$u_p = q_p^{in} + q_p^{out}, \quad p \in \{1, 2, \dots, P\}. \quad (12)$$

In the rest of the paper, *traffic* and *usage* of a participant will be used interchangeably to refer to u_p given by Equation (12). Note that the fixed cost in Equation (11), the first summand, is the same for all participants (i.e. it corresponds to the membership fee). The allocation in Equation (11) is *budget-balanced* (i.e. $\sum g_p = F + S = c_T$) and possesses the *anonymity property* (i.e. the labelling of the participants does not play a role in the assignment).

8.2 Shapley value

The Shapley value is a solution concept in cooperative game theory that fairly and uniquely solves the problem of the distribution of surplus (or joint-cost allocation) among players when considering the worth (or cost) of each coalition [17]. Today, the Shapley value is perhaps the most commonly used method to allocate the costs in cost-sharing games [14] because it is budget-balanced and guarantees equilibrium in any game, regardless of its parameters [10]. We compute the Shapley value (see Appendix A) using Equation (8) as a characteristic function.

8.3 Optimal γ

Let s_p be the cost assigned to participant p by the Shapley value. We call γ^* the value of γ in Equation (11) that minimises the mean square error with Shapley values (i.e. $\min_{\gamma} \sum (g_p - s_p)^2$). Computing $\partial \sum (g_p - s_p)^2 / \partial \gamma = 0$, it can be easily found that a unique critical point exists and has a positive second derivative; thus, it is a global minimum. The value is given by the following:

$$\gamma^* = \frac{\sum_{p=1}^P (s_p / c_T - \eta_p) (1/P - \eta_p)}{\sum_{p=1}^P (1/P - \eta_p)^2}, \quad (13)$$

where $\eta_p = u_p / \sum u_p$, $p \in \{1, 2, \dots, P\}$, $0 \leq \eta_p \leq 1$, $\sum \eta_p = 1$. Thus, η_p represents the relative system usage of participant p . Note that, for an arbitrary mapping between relative usage η_p and Shapley values s_p , γ^* , given by Equation (13), is not bounded (i.e. $\gamma^* \in [-\infty, \infty]$). However, it would not be reasonable to take γ in Equation (11) outside the interval $[0, 1]$. What this result shows is that having $\gamma^* \in [0, 1]$, Shapley values might be approximated well by Equation (11), taking $\gamma = \gamma^*$, while Equation (11) might not approximate Shapley values well when $\gamma^* \notin [0, 1]$. We see that the condition $\gamma^* \in [0, 1]$ holds for our dataset, and we believe that this happens in most practical cases. Furthermore, when $\gamma^* \in [0, 1]$, its value can be used as an indicator of how balanced the system is. If the system is very unbalanced, for example such that a single participant \hat{p} consumes all resources (i.e. $\eta_{\hat{p}} \rightarrow 1$, $\eta_{p \neq \hat{p}} \rightarrow 0$, and thus, $s_{\hat{p}} \rightarrow c_T$, $s_{p \neq \hat{p}} \rightarrow 0$), then from Equation (13) we have the following:

$$\lim_{\substack{\eta_{\hat{p}} \rightarrow 1, \eta_{p \neq \hat{p}} \rightarrow 0 \\ s_{\hat{p}} \rightarrow c_T, s_{p \neq \hat{p}} \rightarrow 0}} \gamma^* \rightarrow 0. \quad (14)$$

Alternatively, if the system is balanced, such that $\eta_p \rightarrow 1/P \forall p$ (i.e. all participants consume the same and, thus, $s_p \rightarrow c_T/P \forall p$), then from Equation (13), we have the following:

$$\lim_{\substack{\eta_p \rightarrow 1/P \\ s_p \rightarrow c_T/P}} \gamma^* \rightarrow 1. \quad (15)$$

Therefore, we conclude that $\gamma^* \rightarrow 1$ for a system where participants equally consume the resources, while $\gamma^* \rightarrow 0$ for a system where most resources are consumed by a reduced number of participants. For our dataset, we obtain $\gamma^* = 0.43$. Thus, we can expect that Shapley values can be approximated well by Equation (11), taking $\gamma = \gamma^*$. We will confirm this with the numerical results explained in the following. Moreover, $\gamma^* = 0.43$ is far from 1, showing a rather unbalanced system, as we have seen in Section 6.1 (the four participants with the highest usage account for 77.12% of the traffic).

Figure 6 (a) shows the cost assignments according to the two previous cost-sharing mechanisms, the Shapley value and Guifi.net, with several values of γ in the last case. All of the distributions are *individually rational* because, in any case, a participant pays more when not cooperating with anyone else (*standalone* curve). Figure 6 (b) shows the normalised values of Figure 6 (a) resulting from dividing the assigned costs by the means of the inbound and outbound traffic, giving the cost in €/Gbps. The €/Gbps is a metric commonly used by ISPs to assess alternatives. The curves of this figure are not strictly monotonically decreasing due to the differences between the metric used in the cost-sharing calculations, u_p , and the metric used for the normalisation, $m(p)$. Figure 6 (c) shows the cost gain, defined as the standalone cost over the cost incurred inside Guifi.net (i.e. the cost depicted in Figure 6 (a)).

Interestingly, Figure 6 reveals that for $\gamma = 0.4$ a striking coincidence exists between costs computed using the Guifi.net method and the Shapley value. Indeed, recall that $\gamma^* = 0.43$. The fixed cost per participant resulting from Equation (11), $\gamma c_T/P$, is 142.37 € for γ^* . This is rather close to the membership fee of 154.00 € used in Guifi.net, which corresponds to $\gamma = 0.47$. In other words, without being aware of it, Guifi.net participants are contributing an assigned cost that is very close to the Shapley value.

8.4 Meaning of γ

Note that, from Equation (10), in one extreme, by setting $\gamma = 0$, we assign the whole cost to the shared part, thus proportionally dividing the cost according to each participant's traffic. At the other extreme, by setting $\gamma = 1$, we assign the whole cost to the fixed part, thus equally weighting the total cost among all participants. As shown in Figure 6 (a), the assignment line varies from a curve with the highest positive slope when $\gamma = 0$ to a flat line when $\gamma = 1$. Therefore, for the participants on the left (with low traffic) the lower the γ is, the lower the cost they pay. While the opposite holds for the participants on the right, with high traffic, the lower the γ is, the higher the cost they pay. The boundary between these regions is given by the point where the cost does not change with γ . Therefore, this point can be computed by $\partial g_p / \partial \gamma = 0$. Using Equation (11), we find that the boundary is given by the following mean usage:

$$\langle u \rangle = \frac{1}{P} \sum_{i=1}^P u_p. \quad (16)$$

This is a logical result because users with a lower usage than the mean, $u_p < \langle u \rangle$, prefer $\gamma = 0$; thus, every participant is paying for his or her usage. On the other hand, users with higher usage than the mean, $u_p > \langle u \rangle$, prefer $\gamma = 1$, thus dividing their cost among all participants. The optimal γ computed before using the Shapley value is a good compromise between these opposite interests.

9 Sensitivity to participants with high usage

The participants with the highest bandwidth consumption, those that contribute the most in absolute terms, might be tempted to leave Guifi.net and build their own coalition, or at least to threaten to form one to impose cost-sharing rules that are more favourable to their interests, for instance, a higher γ .

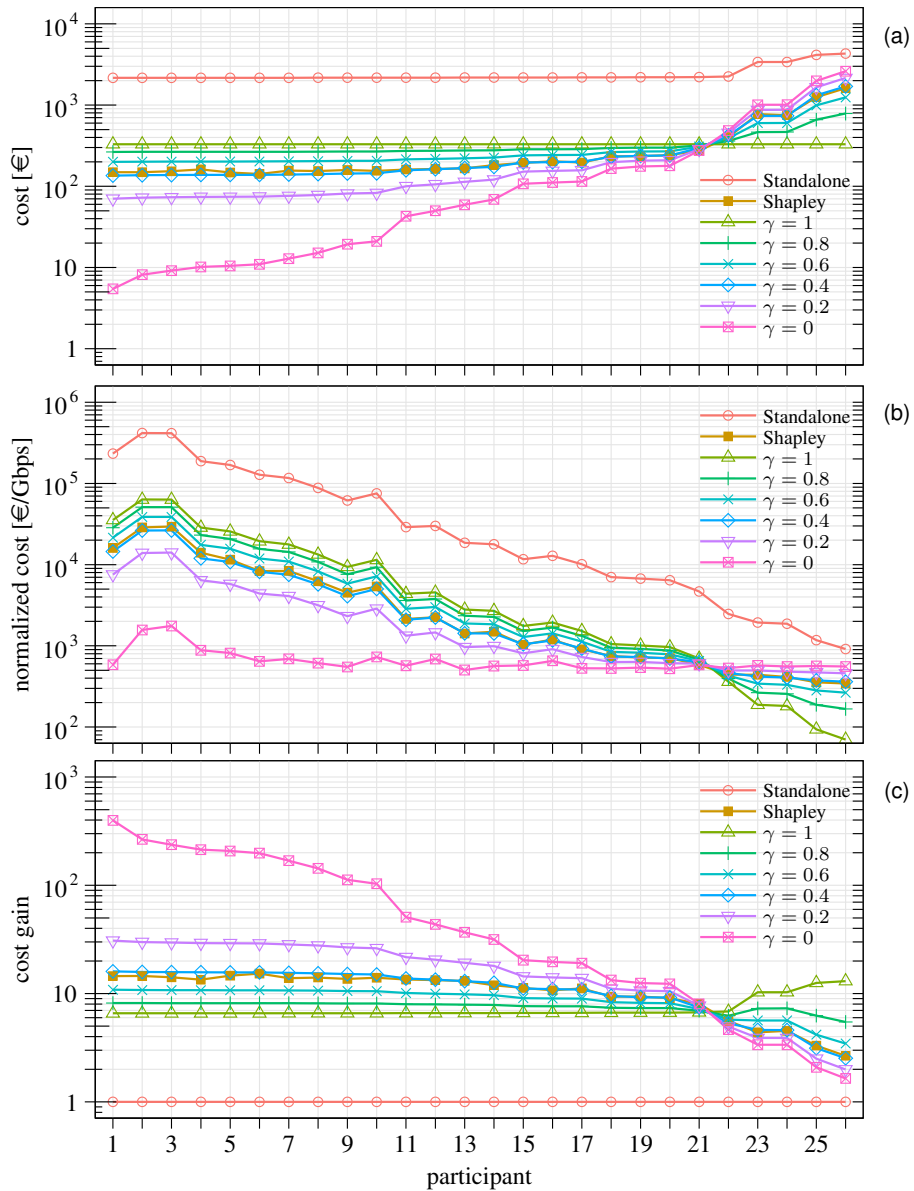


Figure 6: Cost sharing: (a) absolute value, (b) normalised to the traffic mean, and (c) cost gain.

To elucidate whether this reasoning is well-founded, in this section, we investigate the increase in costs that participants would incur upon the largest contributors leaving the current coalition.

Let S be the set of participants leaving Guifi.net and \bar{S} be the remaining set, $\bar{S} \cup S = \{1, 2, \dots, P\}$. Recall that we assume the participants are ordered from the least (participant 1) to the most (participant 26) bandwidth usage (see Eq. (12)). Figure 7 shows the Shapley value assigned to a selected group of participants, including the participants having the least and most usage, s_1 and s_{26} , respectively. The figure also shows the total cost assigned to the participants in S and \bar{S} , $c(S) = \sum_{p \in S} s_p$ and $c(\bar{S}) = \sum_{p \in \bar{S}} s_p$, respectively (see Eq.(8)), and its sum $c(S) + c(\bar{S})$. The abscissa corresponds to the cardinality of the subset S , $|S|$. For instance, for $|S| = 0$, all participants are inside \bar{S} , and for $|S| = 1$, participant 26 has already left the Guifi.net coalition (i.e. $S = \{26\}$). For $|S| = 2$, $S = \{25, 26\}$, and so on. Note that for $|S| = 1$, s_{26} corresponds to the standalone cost of 26, already depicted in Figure 6 (a). The figure shows that between s_{15} and s_{10} , a threshold exists beyond which joining the new coalition is more beneficial than staying in Guifi.net. The figure also shows the local minimums in the staying curves (thin segments) and local maximums in the leaving curves (thick segments). These are because the NSP and IXP cost components of the cost function are step functions. Hence, the unit price functions are not strictly decreasing, which is the condition for economies of scale to take place. This creates intervals with local optima that discourage traffic growth and, thus,

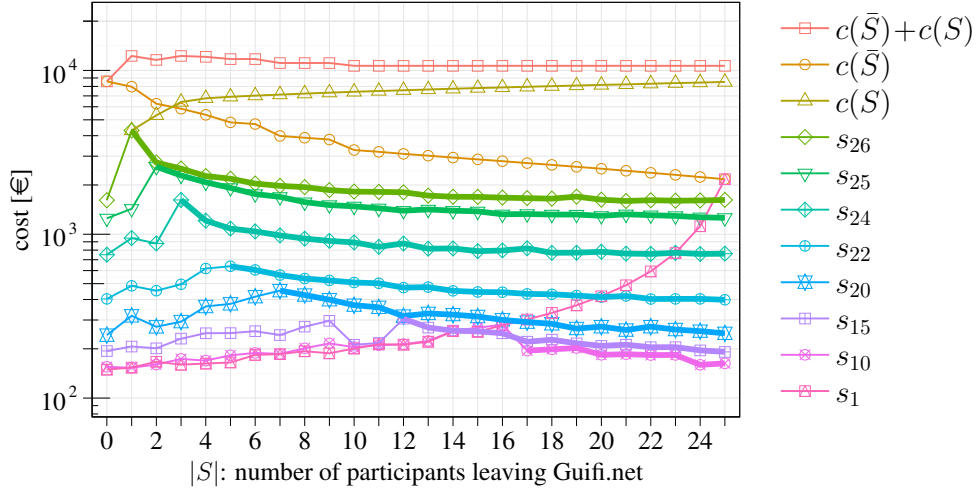


Figure 7: Shapley value of a selected group of participants (1 and 26 are the participants exchanging the least and most external traffic, respectively) upon a subset S of the highest-usage participants leaving Guifi.net, and \bar{S} staying. Segments are thicker after the participant has left Guifi.net.

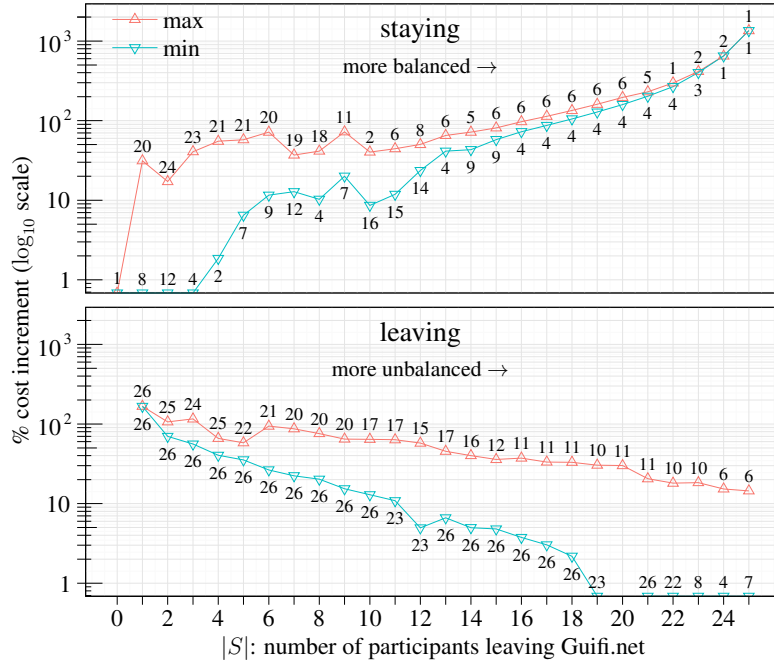


Figure 8: Maximum and minimum percentage cost increment over all participants in \bar{S} (staying in Guifi.net) and S (leaving Guifi.net) corresponding to Figure 7 in log scale.

cooperation. For instance, when participant 25 leaves, the costs incurred by most of the rest who stay decrease, and the same happens for participant 9. Nevertheless, from a general perspective, these local reductions of savings are, by far, compensated for by the region of savings that the greater order of magnitude of products entails (e.g. when moving from combinations of 1 Gbps links to a 10 Gbps setting).

Figure 8 shows the minimum and maximum increments in log scale of the Shapley value for both the participants staying in Guifi.net (top) and those leaving Guifi.net (bottom) with respect to the Shapley value before splitting into the staying and leaving sets. That is, let $r_p(k)$ be the relative cost increment of participant $p \in k$, $k \in \{S, \bar{S}\}$, and $s_p(k)$, $k \in \{S, \bar{S}, A_P\}$ be the Shapley value assigned to participant $p \in k$, where $A_P = \{1, 2, \dots, P\}$ is the set before splitting. Then, we obtain the following:

$$\max_{p \in k} (r_p(k)) = \max_{p \in k} \left(\frac{s_p(A_P) - s_p(k)}{s_p(A_P)} \right), \quad k \in \{S, \bar{S}\}, \quad (17)$$

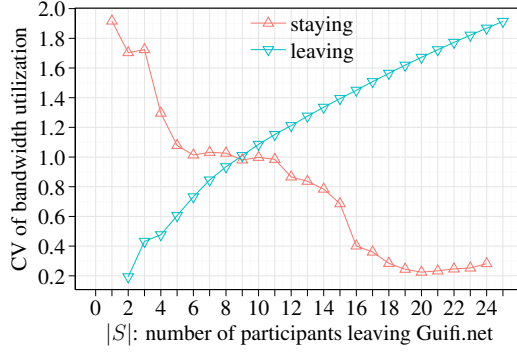


Figure 9: Coefficient of variation (CV) of the participants’ bandwidth usage, u_p , in each staying \bar{S} and leaving S sets. The lower the CV is, the more *balanced* the sets are.

and likewise for the minimum. The labels next to the points indicate which participants reach the max and min cost increments.

Figure 8 shows that the Shapley value increments are very different among the participants. For instance, when $S = \{26\}$ and $|S| = 1$, we find that $\max(r_p(\bar{S})) = 31.26\%$ (reached by participant 20), and $\min(r_p(\bar{S})) = -4.33\%$ (reached by participant 8). Therefore, participant 8 is slightly favoured by the departure of participant 26, while participant 20 has a significant increment in the assigned cost. When more participants with higher usage leave Guifi.net, the set of staying participants are more balanced, and the ratio between the maximum and minimum is lower (top of Figure 8). Recall that the ordinates are in log scale, and we observe that $\log(\max) - \log(\min) = \log(\max / \min)$ diminishes up to 0 at $|S| = 25$, when there is a single participant in \bar{S} . At this point, Figure 8 shows that the cost increment for the single participant in \bar{S} increases more than 1000%. It becomes apparent that the cost would not be affordable for a single participant, nor even for a small group of participants with low usage.

The opposite occurs for the set of leaving participants (bottom of Figure 8). When more participants join the set, the \max / \min ratio is more unbalanced and higher. To assess how balanced these sets are, Figure 9 shows the coefficient of variation (CV) of the participants’ bandwidth usage, measured using Equation (12). Figure 9 confirms that the CV of the staying set tends to 0 (more balanced) as participants leave the set. Alternatively, the CV of the leaving set increases (more unbalanced) as new participants join. Thus, we conclude that, in general, the more unbalanced the set becomes when a new participant joins the set, the greater the ratio will be between the maximum and minimum cost increments that the participants of the perturbed set will have. We can guess that the same occurs when participants leave the set.

If we assume that the participants leave Guifi.net from the lowest usage (participant 1) to the highest usage (participant 26), we would obtain Figure 8 moving from right to left and with the staying and leaving sets being those at the bottom and top, respectively. For instance, if participant 1 leaves Guifi.net, Figure 8 at the bottom shows that participant 6, a participant with low usage, would have an assigned cost increment of 14.44%. This is a remarkable increment, given that the participant leaving is the one with the least usage, and, thus, with a low assigned cost. Furthermore, participant 6 would have a cost increment of -1.15% if participant 26 left instead of 1. We thus conclude that some participants are more sensitive to departures of low-usage participants rather than high-usage participants.

It is also interesting that, in set \bar{S} (staying), the participants with a higher usage are those with higher cost increments, while the opposite occurs in set S (leaving). This is a logical result because, when a participant with high usage leaves Guifi.net, the other participants with high usage must afford the cost that their need for greater resources requires, which had been shared with the participant who left. The opposite occurs for the set S , where the leaving participant joins. Thus, we conclude that, when a participant with high usage joins the set, the participants in the perturbed set that will have the least assigned cost increment will be those of the same type (high usage). Alternatively, if a participant

with high usage leaves the set, the participants of the same type will have the highest cost increment.

Yet again, because the unit price functions of some cost components are not strictly decreasing, on some occasions, the traffic reduction through the reduction of the number of participants is beneficial for some of those who stay, reaching up to 10% in savings between s_9 and s_{10} . Nonetheless, Figure 8 as a whole confirms that the dominant strategy is to be part of the grand coalition.

10 Approximation of the Shapley value using Guifi.net's cost sharing

In Section 8, we found that Guifi.net's cost sharing is close to the cost assignments obtained using the Shapley value. This is a significant result because the Shapley value is difficult to compute for numerous participants. In this section, we analyse how well the Shapley value is approximated using Guifi.net's cost-sharing mechanism given by Equation (11). We investigate the approximation for all sets S (participants leaving Guifi.net) and \bar{S} (participants staying in Guifi.net) analysed before. Note that these sets form coalitions of participants with very different characteristics, as shown by the high variation of the usage CV in Figure 9 (ranging approximately from 0 to 2). Thus, our analysis will compare the approximation for a wide range of situations.

For each S and \bar{S} sets of participants, we consider two cases: (i) when Guifi.net's cost sharing uses γ^* given by Equation (13), and (ii) when Guifi.net's cost sharing uses a fixed membership fee. We denote these cases by γ_k^* and $\hat{\gamma}_k$, $k \in \{S, \bar{S}\}$, respectively. Case (i) gives us the best approximation of the Shapley value (with the minimum mean square error) that can be obtained using Guifi.net's cost sharing. Case (ii) is more practical. Note that computing γ_k^* requires knowing the Shapley values of the participants in the set $k \in \{S, \bar{S}\}$. In practice, it is likely that Guifi.net's cost sharing is used, fixing a reasonable membership fee. The fee can be fixed with different criteria, for instance, such that participants with a low budget can join Guifi.net. In our analysis, we will consider the membership fee currently used in Guifi.net (i.e. 154 €).

Figure 10 shows the value of γ_k^* , $k \in \{S, \bar{S}\}$ given by Equation (13). The curves labelled with $\hat{\gamma}_k$ show the value of γ_k , $k \in \{S, \bar{S}\}$ corresponding to the fixed fee of 154 €, that is:

$$\hat{\gamma}_k = \frac{154 |k|}{c(k)}, \quad k \in \{S, \bar{S}\}, \quad (18)$$

where $|k|$ and $c(k)$ are the number of participants and total cost (Eq. (8)), respectively, of the sets $k \in \{S, \bar{S}\}$.

To assess the approximation, Figure 11 shows the maximum relative error (MRE) in percentage between Guifi.net's cost sharing and the Shapley value assigned to any participant using γ_k^* and $\hat{\gamma}_k$:

$$\text{MRE}(\gamma) = \max_{\forall p} \left(\frac{|g_{p,k}(\gamma) - s_{p,k}|}{s_{p,k}} \right), \quad \gamma \in \{\gamma_k^*, \hat{\gamma}_k\}, \quad k \in \{S, \bar{S}\}, \quad p \in k, \quad (19)$$

where $g_{p,k}$ and $s_{p,k}$ are the Guifi.net's cost sharing (Eq. (11)) and Shapley value, respectively, for each set $k \in \{S, \bar{S}\}$ and for each participant $p \in k$.

We observe the following:

- Figure 10 shows that, in most cases, we obtain $\gamma_k^* \in [0, 1]$, $k \in \{S, \bar{S}\}$. Thus, at least for the cost model derived in this paper, Shapley values might be approximated using Guifi.net's cost sharing using γ^* .
- Figure 11 shows that, in all cases, the MRE corresponding to γ^* , $k \in \{S, \bar{S}\}$ is lower than 20%. Thus, confirming that Shapley values are reasonably well approximated by Guifi.net's cost sharing using γ^* . Furthermore, because γ^* minimises the mean square error, the relative

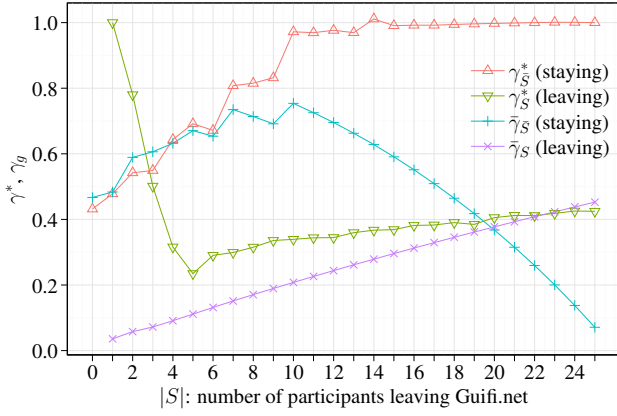


Figure 10: γ_k^* and $\hat{\gamma}_k$ for each set k : S (leaving Guifi.net) and \bar{S} (staying in Guifi.net).

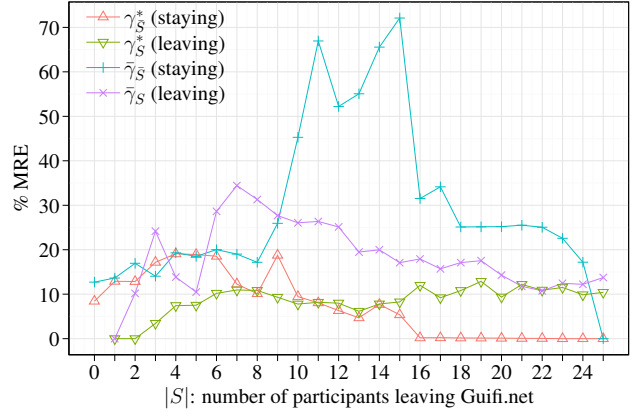


Figure 11: Maximum relative error (MRE) between Guifi.net's cost sharing and the Shapley value.

error is expected to be much lower for participants with higher costs. Indeed, we found that the participant with the highest assigned cost, participant 26, has a lower relative error than 4.14% in all scenarios. This is important because participants with the highest cost assignments are those that both the most about the fairness of the assigned cost.

- In Section 8.3, we derived that, if the system is balanced, $\gamma^* \rightarrow 1$. The *staying* subset becomes more balanced each time as participants with higher traffic leave (see Figure 9). Figure 10 shows that $\gamma_{\bar{S}}^* \rightarrow 1$, as expected.
- In Section 8.3, we found that, if the system is unbalanced, $\gamma^* \rightarrow 0$. The *leaving* subset rapidly becomes unbalanced as participants with lower traffic join the group (see Figure 9). Accordingly, Figure 10 shows that γ_S^* decreases and remains low (less than 0.5) as expected.
- Figure 11 shows that the MRE corresponding to $\hat{\gamma}_k$, $k \in \{S, \bar{S}\}$ is below 20% in most cases. For instance, for $|S| = 0$ (when all participants are within Guifi.net), the maximum error is only 12.72%. Figure 11 shows that the MRE corresponding to $\hat{\gamma}_{\bar{S}}$ increases beyond 50% when a considerable number of participants with higher traffic leave Guifi.net, and the group still has more than 10 participants. This is motivated by the fact that when all participants have lower traffic usage, the *other cost* dominates the *shared cost* in Equation (11). Consequently, the Shapley value equally divides costs among all participants. That is, the membership fee should cover the overall cost, hence the higher MRE($\hat{\gamma}_{\bar{S}}$). In all other cases, even if γ_k^* and $\hat{\gamma}_k$, $k \in \{S, \bar{S}\}$ are significantly different, the MRE corresponding to $\hat{\gamma}_k$ is lower than 40%. We conclude that, for a reasonable value of the membership fee, Guifi.net's cost sharing is not expected to be significantly different from the Shapley value (MRE less than 40%).

11 Tuning Guifi.net's cost sharing

Although the Shapley value can be considered a sensible fair scheme from an economic point of view, it has some drawbacks when used directly as a cost assignment mechanism in Guifi.net or other similar scenarios. Possibly the most important drawback is that the Shapley value is difficult to compute because it needs to formulate the characteristic function and use numerical methods, such as Monte Carlo simulations. Thus, many participants would possibly not be able to compute or estimate the assigned cost and validate whether it agrees with the Shapley value.

On the contrary, the simple Guifi.net cost sharing is easy to understand and easy to estimate. Participants know that they are assigned a membership fee and a variable cost that is proportional to their bandwidth usage. Thus, participants with low usage can estimate that the assigned cost will be close

to the membership fee. Possibly, these participants do not care if the cost they are assigned has a relative error of 20% with the Shapley value, as long it is not far from the membership fee they knew beforehand. Furthermore, Guifi.net’s cost sharing is easily computed using the average traffic, which is a measure available to all participants. Thus, they can easily validate the cost they are assigned.

As seen in Section 8, participants with higher usage in Guifi.net afford costs that are orders of magnitude higher than those of the participants with low bandwidth usage. Such participants are concerned about the cost they are assigned and the fairness of the amount of the membership fee. One convincing justification to them can be the Shapley value. That is, if the cost they are assigned is close to the Shapley value, it is a sound argument to convince them that the cost assignment is fair.

We have seen in Section 10 that, when using γ^* in Guifi.net’s cost sharing, the MRE with the Shapley value is expected to be less than 20%. Because γ^* minimises the mean square error, for participants with high assigned costs, the relative error with the Shapley value is even lower than 20%. Indeed, we found that the participant with the highest assigned cost, participant 26, has a relative error that is lower than 4.14% in all analysed scenarios in Section 10. Therefore, considering that traffic patterns are not much different in consecutive months, we propose the following modification to Guifi.net’s cost-sharing approach. Every time the membership fee is to be updated, the following should be performed:

1. Compute γ^* using Equation (13) and the dataset of the previous month. We assume that we obtain $\gamma^* \in [0, 1]$.
2. Compute the membership fee M corresponding to γ^* :

$$M = \frac{\gamma^* c_T}{P}. \quad (20)$$

Thus, Guifi.net cost sharing will be used with the membership fee of M . Participants with high usage who are worried about the fairness of the cost assignment could compute the Shapley value and validate that it has a small deviation from the assigned cost.

12 Arguments against ceasing cooperation

Cooperation is an inherent part of Guifi.net history. It is only through collaboration and aggregation of demand that small players can access competitive solutions, such as the architecture presented in Section 3 and the associated services. All participants started from scratch; thus, they directly benefited from the joint efforts to overcome the telecommunications’ sector entry barriers. The forms of cooperation have evolved over time from the initial informal agreements around small and rather isolated projects to the current well-structured and legally binding body of normative agreements to meet new external challenges and the changes in interests, expectations, and objectives of the participants. This transformation process, which involved long discussions, agreements, and concessions among the participants, has had to cope with internal tensions recurrently jeopardising the cooperation. The threat of ceasing cooperation by one or more members has frequently been one of the causes of this tension or has at least been used as a means of enforcing internal rule changes.

In this regard, our findings not only provide arguments in favour of staying within the guifi.net coalition but also to focus the efforts towards integrating new participants and increasing the resource usage. As Figure 7 shows, leaving would entail the broadly accepted fact that considerable extra costs would result for the remaining participants and for those who leave, which is something much less well known or at least less publicly acknowledged by those promoting the schism. Figure 8 illustrates that the losses greatly exceed the meagre benefits of the few participants when others leave. Conversely, both figures together, read from right to left, underline the benefits of the increasing number of participants and amount of usage consumption. Lastly, to obtain an approximation of the

order of magnitude of the losses due to the cease of cooperation, we must consider that the external traffic case we analyse in this paper is just a minor fraction of the total activity volume in Guifi.net for these ISPs.

13 Conclusion

In this paper, we analysed an economic case study of collaborative networking. More specifically, we addressed the economy of Guifi.net, the largest community network in the world. Guifi.net started in 2004 as a non-profit community network with the objective to provide high bandwidth access to the Internet in rural areas. In contrast to many similar experiences that stagnated, Guifi.net has had enormous growth. At the time of writing, Guifi.net encompasses a large number of users grouped into what we call *participants* in this paper. These participants range from groups of enthusiasts willing to operate their own network in a free open and neutral approach, which is the spirit that originated Guifi.net, to actual ISPs with commercial objectives and thousands of customers. These types of participants were able to build their profitable businesses due to the knowledge and cost-sharing principles that govern Guifi.net. The inclusion of such ISPs has definitively contributed to the success of Guifi.net. Of course, keeping participants with these diverse interests together is challenging.

In this paper, we explained and numerically assessed the costs related to Guifi.net's external traffic. We built a mathematical economic model that describes these costs. That is, we derived a model that estimates the costs of an arbitrary set of participants in terms of their traffic parameters. We provided a detailed description of the costs to be considered in the model: IXP and carrier costs, management costs, electricity costs, etc. To the best of our knowledge, this is the first paper providing such a detailed economic model of the external traffic networking costs.

With the economic model, we analysed the cost-sharing approach used inside Guifi.net. Such an approach uses a simple method consisting of equally dividing a part of the total cost among the participants, while proportionally dividing the other part according to each participant's bandwidth usage. To assess its fairness, we compared it with the Shapley value. Surprisingly, we found that, even if Guifi.net users are not aware of it, Guifi.net's cost-sharing model assigns a cost that is very close to the Shapley value. Furthermore, we found that the simple sharing approach of Guifi.net can approximate the Shapley value in a wide range of cases. We believe that this is a significant contribution because the Shapley value for a large number of participants is difficult to compute.

Our numerical results reveal the costs savings that collaborative networking inside Guifi.net provides to the participants. We numerically assessed the cost increments that the highest-usage participants would incur upon leaving Guifi.net. We found, for instance, that the four highest-usage participants account for 77.12% of the traffic. We estimate that, upon these participants leaving Guifi.net, the staying and leaving groups of participants would have increased overall external traffic costs by 28.11% and 54.01%, respectively. In summary, we hope that our results will help all participants to better appreciate the advantages of staying with Guifi.net.

Acknowledgments

The authors want to express gratitude for the valuable datasets, insights, and comments received from the Guifi.net Foundation. This work was supported by Spanish grant TIN2016-77836-C2-2-R and Generalitat de Catalunya through 2017-SGR-990 and 2016-DI-006.

A Shapley value

The Shapley value is a well-known concept in coalitional game theory. For the sake of completeness in this appendix, we provide a brief review of the Shapley value adapted to our problem, using the notation defined in previous sections. The selected literature about the Shapley value can be found in [15].

Our coalitional game consists of the pair (A_P, c) , where

- $A_P = \{1, 2, \dots, P\}$ is the set of participants (the *grand coalition*), and
- $c : 2^P \rightarrow \mathbb{R}$ associates with each coalition $A \subseteq \{1, 2, \dots, P\}$, a real-valued cost, $c(A)$, given by Equation (8). We assume $c(\emptyset) = 0$.

The Shapley value assigns the cost to each participant:

$$s_p = \frac{1}{P} \sum_{A \subseteq A_P \setminus \{p\}} \frac{c(A \cup \{p\}) - c(A)}{\binom{P-1}{|A|}}, \quad p \in \{1, 2, \dots, P\}, \quad (21)$$

where the summation extends all subsets of A_P not containing participant p (with one of the subsets being the empty set), thus resulting in 2^P summands. Note that the idea behind Equation (21) is assigning the average marginal cost to each participant, that is, the average increment of the cost that each participant adds to every possible coalition where he or she is included.

The Shapley value is the unique assignment that satisfies the following desirable properties:

1. *Budget-balanced*: $\sum_{p=1}^P s_p = c(A_P)$.
2. *Symmetry*: If $\forall A \subset A_P : c(A \cup \{p_1\}) = c(A \cup \{p_2\})$, then $s_{p_1} = s_{p_2}$.
3. *Dummy player*: If $\forall A \subset A_P, p \notin A : c(A \cup \{p\}) - c(A) = c(\{p\})$, then $s_p = c(\{p\})$.
4. *Additivity*: If we combine two coalitional games (A_P, c_1) and (A_P, c_2) , in $(A_P, c_1 + c_2)$, then $s_p(c_1 + c_2) = s_p(c_1) + s_p(c_2)$.

Due to the exponential growth of the summands, computing the Shapley value using Equation (21) is impracticable for large P . In our research, we used the R package [16] to compute Equation (21) up to $P = 10$. For larger values of P , we used a Monte Carlo simulation (see e.g. [12]). We provide our source code in the public repository [18].

References

- [1] Fiona H. Ashmore, John H. Farrington, and Sarah Skerratt. “Superfast Broadband and Rural Community Resilience: Examining the Rural Need for Speed”. In: *Scottish Geographical Journal* 131.3-4 (2015), pp. 265–278.
- [2] Roger Baig et al. “guifi.net, a crowdsourced network infrastructure held in common”. In: *Computer Networks* 90 (2015), pp. 150–165.
- [3] Roger Baig et al. “Making Community Networks Economically Sustainable, the Guifi.Net Experience”. In: *Proceedings of the 2016 Workshop on Global Access to the Internet for All, SIGCOMM 2016*. GAIA ’16. Florianopolis, Brazil: ACM, 2016, pp. 31–36.
- [4] *Catalunya Neutral Internet eXchange*. <http://www.catnix.net/en>.
- [5] *CATNIX tariffs*. <http://www.catnix.net/ca/tarifefes>.

- [6] Xenofontas Dimitropoulos et al. “On the 95-Percentile Billing Method”. In: *Proceedings of the 10th International Conference on Passive and Active Network Measurement*. PAM '09. Seoul, Korea: Springer-Verlag, 2009, pp. 207–216.
- [7] *Equinix Data Center Barcelona*. <https://www.equinix.com/locations/europe-colocation/spain-colocation/barcelona-data-centers/>.
- [8] *Fundació privada per a la xarxa oberta, lliure i neutral guifi.net*. <https://fundacio.guifi.net/>.
- [9] Hernán Galperín. “Localizing Internet infrastructure: Cooperative peering in Latin America”. In: *Telematics and Informatics* 33.2 (2016), pp. 631–640.
- [10] R. Gopalakrishnan, J. R. Marden, and A. Wierman. “Characterizing distribution rules for cost sharing games”. In: *International Conference on NETWORK Games, Control and Optimization (NetGCooP 2011)*. Oct. 2011, pp. 1–4.
- [11] *Guifinet anonymized CSV traffic repository*. <https://gitlab.com/llorenc/guifinet-anonymized-csv.git>.
- [12] Sasan Maleki et al. “Bounding the estimation error of sampling-based Shapley value approximation”. In: *arXiv preprint arXiv:1306.4265* (2013).
- [13] Murtaza Motiwala et al. “Towards a Cost Model for Network Traffic”. In: *SIGCOMM Comput. Commun. Rev.* 42.1 (Jan. 2012), pp. 54–60.
- [14] Robin Pilling, Shi Chung Chang, and Peter B. Luh. “Shapley Value-Based Payment Calculation for Energy Exchange between Micro- and Utility Grids”. In: *Games* 8.4 (Oct. 2017), pp. 1–12.
- [15] Alvin E Roth. *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press, 1988.
- [16] Alejandro Saavedra-Nieves. *GameTheoryAllocation: Tools for Calculating Allocations in Game Theory*. R package version 1.0. 2016.
- [17] Roberto Serrano. *Cooperative games: core and Shapley value*. 2007.
- [18] *Shapley value Monte Carlo simulation with R*. <https://gitlab.com/llorenc/shapley-value>.
- [19] LS Shapley. “A value for n-person games. Contributions to the Theory of Games II. Ed. by HW Kuhn and AW Tucker”. In: *Annals of Mathematics Studies* 28 (1953).
- [20] Rade Stanojevic, Ignacio Castro, and Sergey Gorinsky. “CIPT: Using Tuangou to Reduce IP Transit Costs”. In: *Proceedings of the Seventh Conference on Emerging Networking Experiments and Technologies*. CoNEXT '11. Tokyo Japan: Association for Computing Machinery, 2011, pp. 1–12.
- [21] Rade Stanojevic, Nikolaos Laoutaris, and Pablo Rodriguez. “On economic heavy hitters: shapley value analysis of 95th-percentile pricing”. In: *Internet Measurement Conference*. ACM, 2010, pp. 75–80.
- [22] S. H. Tijs and T. S. H. Driessen. “Game Theory and Cost Allocation Problems”. In: *Management Science* 32.8 (1986), pp. 1015–1028.
- [23] William Waites et al. “RemIX: A Distributed Internet Exchange for Remote and Rural Networks”. In: *Proceedings of the 2016 Workshop on Global Access to the Internet for All*. GAIA '16. Florianopolis, Brazil: ACM, 2016, pp. 25–30.
- [24] Wikipedia contributors. *Burstable billing* — *Wikipedia, The Free Encyclopedia*. 2018. URL: https://en.wikipedia.org/w/index.php?title=Burstable_billing&oldid=844986693 (visited on 07/05/2019).