Network and Revenue Analysis of an Affiliate Marketing Program in the Travel Industry

Lucas L. Rolim[1,2], **Jefferson E. Simões**[3], **Daniel R. Figueiredo**[2]

[1]Hurb.com
Rio de Janeiro, Brazil

[2]Federal University of Rio de Janeiro (UFRJ)
Rio de Janeiro, Brazil

[3]Federal University of the State of Rio de Janeiro (UNIRIO)
Rio de Janeiro, Brazil

`lucas.rolim@hurb.com,jefferson.simoes@uniriotec.br, daniel@cos.ufrj.br`

Abstract. *The widespread adoption of e-commerce over the past two decades has transformed businesses and sparked novel marketing strategies. In affiliate marketing individuals sign up with companies to promote or sell their products in independent venues such as blogs and channels controlled by the affiliate, receiving compensations for their actions. This work analyzes* Clube Hurb, *a real and large affiliate marketing program, considering the affiliate network structure, the revenue generated by affiliates, and their relationship. While the network is largely fragmented (90.7% of the affiliates are isolated) and most affiliates never sell (99.5%), different network and revenue statistics exhibit heavy-tailed behavior and are sometimes correlated. The findings shed light on affiliate marketing dynamics and can drive future studies to improve performance.*

1. Introduction

Electronic commerce (e-commerce) is among the most disruptive services enabled by the Internet with drastic impact on multiple stakeholders such as manufactures, retailers and consumers [Laudon and Traver 2019]. The selling and buying of goods and services online has transformed various markets and continues to grow both in size and scope, in particular with the advancements of online marketplaces and alternative payment systems.

The widespread expansion of e-commerce has also transformed the role of marketing and advertisement. In particular, online social networks (such as Facebook and Instagram) and search engines (such as Google) provide a novel paradigm for marketing and advertisement. Within the context of online social networks, *viral marketing* has emerged as promising and effective technique for product or service advertisements [Ho and Dempsey 2010, Reichstein and Brusch 2019, Watts and Peretti 2007]. A less studied approach is *affiliate marketing*, which is becoming more widely adopted by companies, given its promises of boosting sales [Duffy 2005, Dwivedi et al. 2017, Ivkovic and Milanov 2010].

In affiliate marketing, a person (the affiliate) signs an agreement with a company to promote and/or sell products and services in venues independent from those of the company, such as websites, blogs and online social networks. In return, the company

pays a commission to the affiliate for different kinds of actions such as selling a product or attracting customers to the company's website.

Affiliate marketing can also lead to *affiliate networks*. In order to increase the number of affiliates, a company often compensates an affiliate that brings other affiliates to their program through referrals. This naturally leads to a network (a tree) where nodes are affiliates and the parent of a node is the affiliate that invited that node into the program. Note that the root of the tree is an affiliate that enrolled in the program directly, and not through referral.

An important consideration is the relationship between the affiliate network and the success of the affiliate marketing program. For example, in viral marketing, the social network structure plays a significant role on a marketing campaign, in particular the presence of hubs (i.e., individuals with very large number of connections). In such contexts, the *reproduction number* is often taken as a metric to quantify the success of the campaign [Watts and Peretti 2007, Leskovec et al. 2007, Hinz et al. 2011, Nguyen et al. 2016]. Do affiliate networks also have such properties? What is the relationship between the affiliate network and the revenue generated by the program? How do the topological properties of a node impact his contribution to the network performance? The present work contributes in this direction with an extensive quantitative characterization of a real affiliate marketing program, focusing on the affiliate network and associated revenue.

Clube Hurb is the affiliate marketing program of Hurb.com, the largest Online Travel Agency (OTA) in Brazil and one of the largest in the world, selling around 60 thousand trips per month to hundreds of thousands of travelers. The affiliate program was launched in April 2018 and offered direct enrollment (via ads on Hurb.com, for instance), which was later expanded to include referral by enrolled affiliates. By January 2020, the program had around 186 thousand affiliates. Using this dataset, this present work makes the following contributions:

- Characterization of the affiliate network: the network is a forest (set of disjoint trees) which is characterized in terms of their sizes, depth, and degrees. Interestingly, the vast majority of trees (96.7%) have size 1: affiliates that were not referred into the program and did not refer anyone. However, the tree sizes follow a very unequal, heavy-tailed distribution (details in Section 3).
- Characterization of the affiliate's revenue: affiliates sell products that adds to the revenue of the company which is characterized in terms of sales and value. Interestingly, the vast majority of affiliates (99.5%) never sell, while some have sold more then 400 products. Revenue generated by affiliates is heavily skewed exhibiting a heavy-tailed distribution (details in Section 4).
- Characterization of the relationship between network and revenue. Interestingly, the average revenue generated by an affiliate that has a parent is three times larger. However, the number of children has no correlation with the revenue of the parent (details in Section 5).

The empirical findings presented in this work sheds light on the diverse dynamics within an affiliate marketing program, where various network and revenue characteristics exhibit heavy tailed behavior. This gives rise to two kind of affiliates: ones that join the program, possibly for fun or in the heat of the moment, and do not remain active (the

vast majority), others that join the program to receive compensations (a small minority). Engaging the former into making sales could significantly improve the revenue generated by the marketing program, opening a discussion on how compensation should be assigned to the different actions.

The remainder of this paper is organized as follows. Details of *Clube Hurb* and the dataset analyzed are provided in Section 2. Sections 3 and 4 present the characterization of the network and revenue, respectively. Their relationship is presented in Section 5. Section 6 discusses the related work on affiliate marketing. Finally, a brief conclusion and discussion is provided in Section 7.

2. Clube Hurb and Dataset

Hurb.com is the largest Online Travel Agency (OTA) in Brazil and one of the largest in the world, selling thousands trips per month to hundreds of thousands of travelers. In April 2018, it launched a public affiliate marketing program called *Clube Hurb* to create a novel sales channel, attract new clients, and increase the visibility of the brand. The program offered direct affiliate enrollment (via ads on its website, for instance), and to boost enrollment, Hurb.com advertised its affiliate marketing program on platforms like Google and Facebook. An affiliate is paid a flat commission rate of 6.5% for any product it sells, though special commission rates apply in some occasions (such as Black Friday). To help affiliates sell products, *Clube Hurb* provides a platform for affiliates to create personalized ads that can then be shared on social media or posted on blogs or websites.

Clube Hurb also allows enrollment through referral by current affiliates. In particular, an affiliate can refer the program to others (via ads on its own blog, for example) who can then become affiliates. When an person enrolls through a referral, the referee earns 15% of the commission received by every sale of the new affiliate. Note that this commission is paid only for direct referrals, and not referrals of referrals and so on.

Note that the program rewards an affiliate by the performance of their referrals (and not when a referral joins). Thus, every referral that is converted into an affiliate becomes a potential source of income for the referee. This provides strong incentives for an affiliate to bring in new affiliates. At the same time, direct enrollment bares no cost to the new affiliate, as its commission on sales is not shared by a parent affiliate.

The dataset used in this work was yielded by Hurb.com and collected by one of the authors while working in the company; it considers all affiliates registered in *Clube Hurb* between the program's inception in April 2018 and January 2020. Various kinds of data were made available for this study, including affiliates that joined through referrals, sales and earnings of affiliates, and revenue for the company. Figure 1 shows the total number of affiliates enrolled in the program over time. Note that *Clube Hurb* did not have an exponential growth at first, but its growth rate has increased significantly over the last six months. Figure 2 shows the total revenue generated by the program for the company. Interestingly, total revenue is highly correlated with the total number of affiliates

3. Network Analysis

The process of enrollment in the affiliate program through referrals naturally induce a directed social network. In this network, every affiliate is a node, and a direct edge from
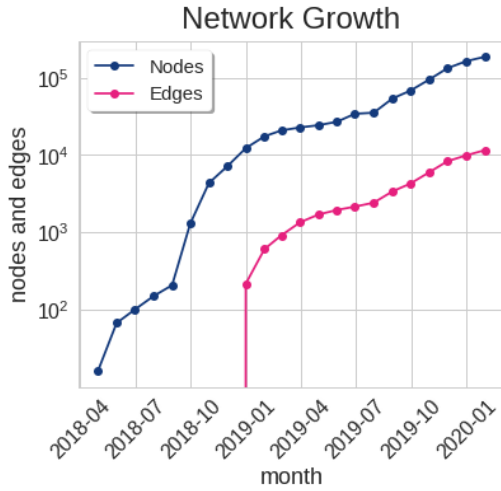
Network Growth

Revenue Growth

Figure 1. Affiliates and network growth over time.
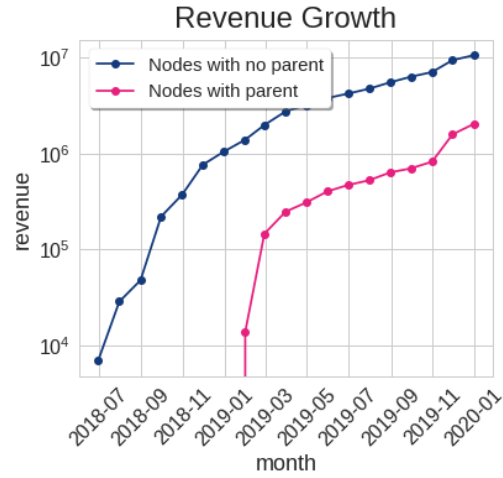
Figure 2. Revenue growth over time (in R$).

u to v exists if v enrolled in the affiliate program through a referral from u. This is called as the affiliate network of *Clube Hurb* and is denoted by the graph $G = (V, E)$ where V is the set of nodes (affiliates) and E the set of directed edges.

Recall that affiliates may join the program directly or through a referral. These two situations result in nodes with in-degree equal to 0 and 1, respectively. Clearly, the network has no cycles, since a node has at most one parent. Therefore, by the nature of referrals, the network is a set of directed trees, i.e., a directed forest[1].

Figure 1 shows the network growth over time, both in terms of nodes and edges. Note that for the first 10 months, the program did not offer enrollment through referrals (and thus no edges could be added). Once referral was allowed, though, the number of edges started increasing in a similar trend as the number of nodes (although at a different scale), indicating an almost constant fraction of new affiliates joining through referrals.

Table 1. General network features (as of January 2020)

Attribute	Value
Number of nodes	186,248
Number of edges	11,536
Number of connected components	174,712
Number of isolated nodes	169,081
Largest connected component	309
2nd largest connected component	174
Fraction of nodes with a parent	6.19%
Fraction of nodes with no child	96.68%
Fraction of nodes with one child	2.65%

Table 1 provides some basic statistics of this network. Interestingly, the vast majority of nodes (96.68%) have out-degree zero, i.e., their referrals have not led to any addi-

---

[1]More precisely, each connected component of the network is an *out-tree* or *arborescence*: a directed tree with edges pointing away from the root.

tional nodes joining the program. Accordingly, the vast majority of nodes (93.81%) have registered on the program directly. Moreover, note that the number of connected components (i.e., trees) is enormous. Most nodes (90.78%) are isolated (both their in-degrees and out-degrees are zero), thus each constituting a connected component by themselves. Such nodes essentially did not participate in referrals: they joined the program directly and their referrals did not lead other nodes to join the program. This is not, however, the only relevant feature of G with respect to the out-degrees. Figure 3 presents the out-degree distribution of G. We notice a heavy tail in this distribution as, despite an average degree of 0.06 and median degree of 0, a non-negligible number of outliers have degree larger than 50. The maximum out-degree in G is 228, 3800 times larger than the average out-degree.
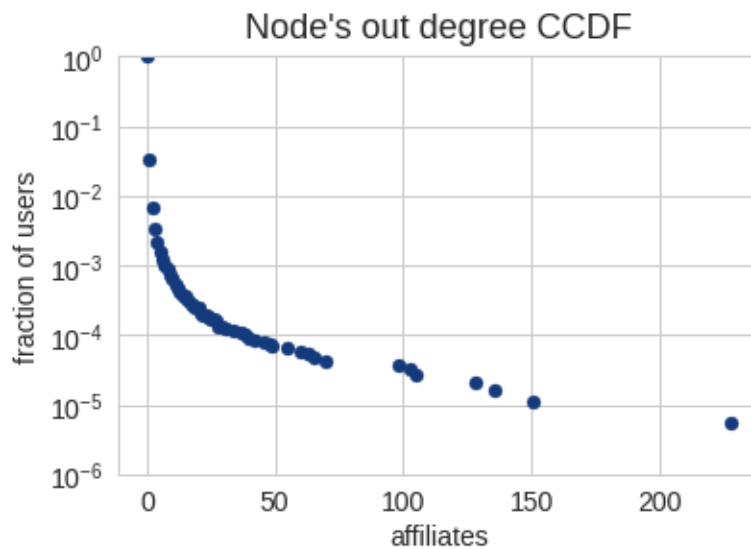

Figure 3. Out-degree distribution of G

This particular process of network formation can be thought of as an epidemic process over a certain population, with membership to the program being "spread" and the affiliate network portraying the spreading process. Through this lens, the average out-degree of the network indicates the expected number of new "infectious" people generated by each existing one [Watts and Peretti 2007], playing the role of *reproduction rate* for this epidemic process. In this particular network, the resulting reproduction rate of 0.06 can be considered extremely low and in the subcritical range.

Figure 4 shows the connected component size distribution. We notice this distribution also displays a heavy tail characteristic: the largest connected component comprises 309 nodes, 78% more than the second-largest connect component (which has 174 nodes), and a particularly high amount since the majority of the connected components has size 1 (isolated nodes). Interestingly, the tree sizes are essentially driven by their most connected nodes. Figure 5 displays the maximum degree, averaged over all connected components of a given size. There is a strong correlation between these two quantities, with near-equality for connected components up to 50 nodes and an almost linear relationship with slope below 1 for larger connected components. This indicates that the tree size is fundamentally a function of its most influential node.
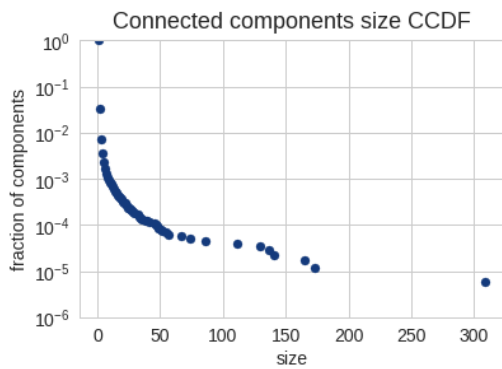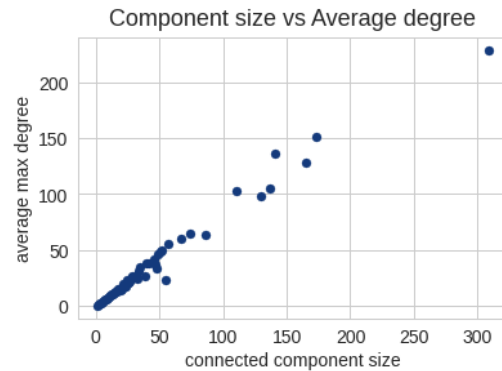
Figure 4. Connected components (tree) size

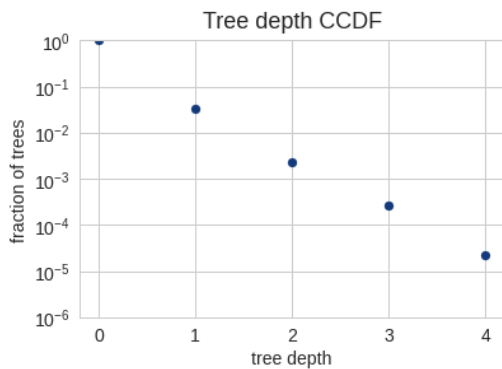**Figure 5. Tree size vs maximum degree**
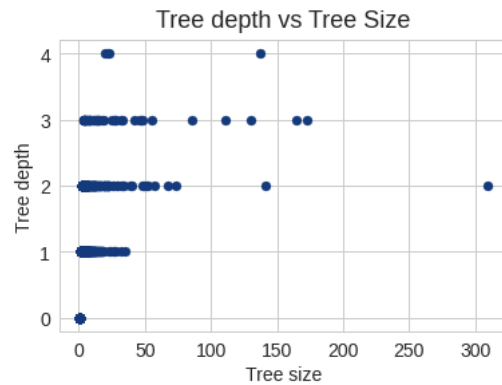

Figure 6. Tree depth

Figure 7. Tree depth vs tree size

As a final purely structural feature, we study the depth of referral cascades. A *referral cascade* occurs when nodes that joined the network through referrals also succeed in recruiting additional users to the network. When a single referral is cascaded from a node that joined directly (the root of a tree), we will say this constitutes a cascade of depth 1, if this referral succeeds in recruiting a new user (and no more) the cascade has depth 2, and so one. Topologically, this is portrayed by directed paths in G, starting at the root of tree with intermediate nodes of the path being responsible for "cascading" referrals, and reflected in the tree depth. Figure 6 shows that nearly all cascades have depth 1, that is, nodes that joined the network through referrals do not succeeded in bringing additional nodes to the network. This is expected given the near equality between connected component sizes and the maximum degree in small components, and the fact that these constitute the vast majority of connected components with size larger than one.

4. Revenue Analysis

The primary purpose of *Clube Hurb* is to generate revenue, and this metric essentially drives strategies and indicates network health. In order to make sales and generate revenue, affiliates can publish links to Hurb.com offers, or generate custom marketing banners that can be embedded in websites to promote the affiliate system. Both links and banners include their referral code, in order to help trace the impact of these actions in terms of marketing and visilibity.

Table 2. Affiliate engagement in activities

Attribute	Value
Fraction of affiliates with pageview:	11.97%
Fraction of affiliates with ads created:	5.74%
Fraction of affiliates that invite other member:	0.67%
Fraction of affiliates with sales:	0.52%
Fraction of affiliates with recurrent sales:	0.27%

Table 2 provides statistics about the activities performed by affiliates. These activities have been sorted in increasing order by difficulty level, with top activities requiring less effort to be performed and/or accomplished. The most fundamental action that a user can perform is to publish weblinks to product offers on its independent channels. This results in pageviews, credited to the affiliate that generated the weblink. Other activities include creating advertisement banners, referrals to the affiliate program, and selling products on Hurb.com. An analysis of the dataset shows that only 11.97% of the affiliates have a pageview in their links, and just 0.52% achieve at least one sale, with half of these users (0.25%) making a single sale.
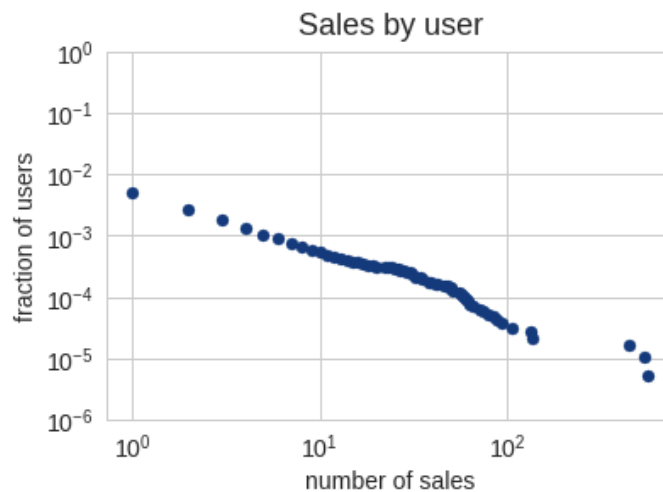

Figure 8. Sales by user

The action of selling products is particularly relevant in terms of generating revenue. Figure 8 shows the distribution of affiliates' sales. Note the heavy tail, potentially a power law, in this distribution: over 99% of affiliates do not sell anything, and a tiny group sells more than 400 items each. Affiliates can have completely different behaviors in terms of sales, thus not all nodes contribute equally to the performance of the affiliate program.

This is evidenced by Lorenz curves for revenue and earnings. The *Lorenz curve* is often used in economics for inequality analysis and depicts the fraction of a resource that is held by the most resourceful individuals; in this case, each point in the curve indicates the fraction of total revenue that is generated by the most productive affiliates in the network (excluding affiliates with no sales). The normalized area between this curve and the straight diagonal line (which represents perfect equality) is known as the *Gini*
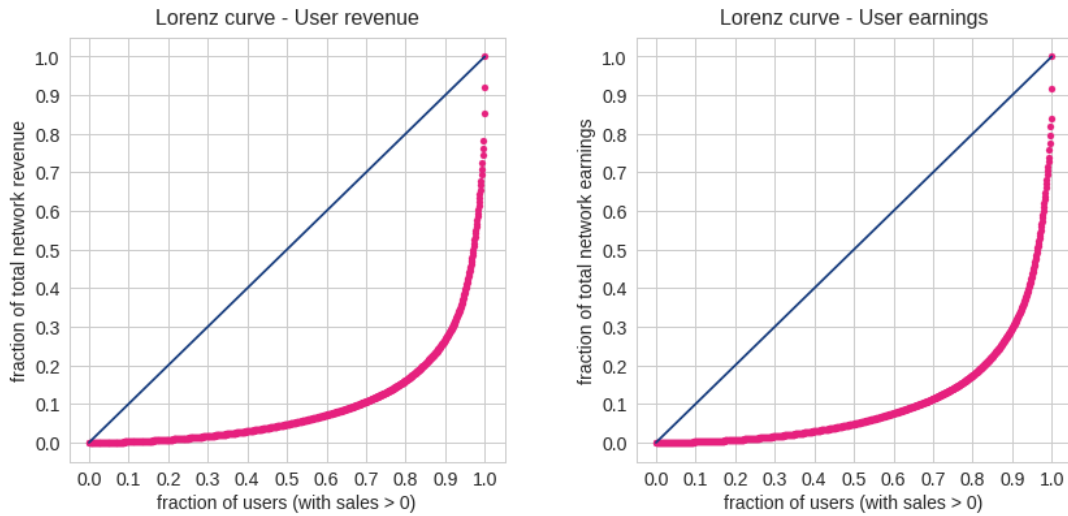
Figure 9. Lorenz curves for revenue and earnings.

coefficient and summarizes the distribution's level of inequality in a scale between 0 and 1 (larger means less equal).

The Lorenz curves for revenue and affiliate earnings are shown in Figure 9 (recall that the earnings of an affiliate comprise the sum of the commission rates applied to each of his sales). The inequality in both curves is notable: for instance, the top 1% affiliates concentrate around 73% of the revenue generated, and a Gini coefficient[2] of around 0.8 for revenue.

5. Network vs. Revenue

While separate analyses of network topology and revenue data reveal important aspects of the affiliate program, there is also important properties that require an analysis of the relationship between the two. More precisely, we would like to find out whether the affiliate network structure has impact on the total revenue generated, other than by helping the system itself grow.

In particular, we would like to evaluate whether having joined the network directly or through a referral has any impact in an affiliate's generated revenue. The average revenue generated by a given affiliate profile constitutes its *lifetime value* or LTV. Table 3 summarizes the LTV for several affiliate profiles. Note that there is an almost two-fold increase in LTV from nodes without parents to nodes with parents. At first sight, it would seem that having a parent is a performance-boosting factor.

Indeed, looking only at users with at least one sale, having a parent has no qualitative influence, with LTV being essentially equivalent between nodes with and without parents. Note that, in both cases, the LTV is nearly 30 times that of the general population. This leads us to conclude that having parent actually increases the likelihood that an affiliate makes a sale, and this, in turn, leads to a much higher individual generation of revenue.

---

[2]For the sake of comparison, income distribution data from the World Bank indicates a Gini coefficient for income of around 0.6 in the most income-unequal countries.

Table 3. Average revenue per affiliate (LTV)

Affiliate group	LTV (R$)
Program population	70.59
Affiliates with parents	175.79
Affiliates without parents	60.32
Affiliates with parents, 1+ sales	1925.84
Affiliates without parents, 1+ sales	2018.03

Finally, we explore the relationship between a node's earnings and its out-degree. Intuitively, we expect these quantities to be positively correlated with the node's level of engagement in the program, thus being themselves positively correlated. Yet, Figure 10 shows there is no meaningful correlation between them.
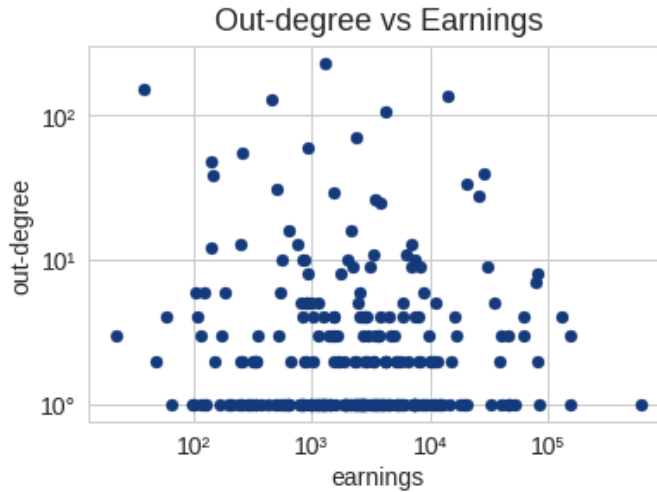

Figure 10. Affiliate earnings vs out-degree (childs). Pearson correlation of 0.03.

6. Related Work

The widespread expansion of e-commerce world wide has profoundly transformed different market segments among retailers and consumers, to name a couple. Along with it, marketing and advertisement has also been transformed over the past couple of decades. New marketing paradigms such as viral marketing and affiliate marketing have emerged and proven to be effective, playing an important role in the online ecosystem [Ho and Dempsey 2010, Reichstein and Brusch 2019, Watts and Peretti 2007, Duffy 2005, Ivkovic and Milanov 2010, Gregori et al. 2014].

Intuitively, viral marketing is the online version of word-of-mouth, where a person tells another something it believes to be interesting. Over the past two decades, the online community has grown exponentially along with the number of channels for interaction (online social networks, blogs, social media, etc). Thus, viral marketing has been hailed to have an enormous potential, yet to be fully captured [Leskovec et al. 2007, Ho and Dempsey 2010, Reichstein and Brusch 2019, Goel et al. 2012]. Problems such as how to start a successful viral marketing campaign has attracted much attention from academia and industry [Hinz et al. 2011, Nguyen et al. 2016].

In affiliate marketing individuals sign an agreement with a company to promote and sell its products and services in independent platforms, such as websites, blogs and channels owned or controlled by the affiliate [Duffy 2005, Ivkovic and Milanov 2010, Gregori et al. 2014]. In this scenario, the affiliate receives direct compensation from the company for its actions, be it sell products, attract customers to the company's website, or refer new affiliates. While affiliate marketing is a fast growing trend among companies, it is much less studied both in academia and industry when compared to viral marketing [Suryanarayana et al. 2019, Dwivedi et al. 2017].

A fundamental aspect of affiliate marketing is determining the compensation value for the various actions of its affiliates as well as revealing a company's internal information to the affiliates (in order to promote/sell products and services). Recent work has shown (theoretically) that compensation values and information disclosure are critical for the success of affiliate marketing as well as being very sensitive [Suryanarayana et al. 2019]. If compensations are not carefully determined, the returns of the affiliate marketing program can be significantly hindered.

Consumer trust on affiliates has also been recently shown to be an important aspect that drives the success of individual affiliates, as well as the affiliate marketing program [Daniele et al. 2009, Gregori et al. 2014]. In particular, by promoting its affiliates the company can boost their trust and improve returns. In the context of tourism, this has been investigated quantitatively through small scale surveys [Daniele et al. 2009].

Affiliate marketing has also been victim of fraud where the attacker receives commission from sales and other actions of common customers (customers not engaged by any affiliate) [Snyder and Kanich 2015, Chachra et al. 2015]. Such attack is possible due to the *cookie stuffing* vulnerability, where the browser of a user is contaminated with a cookie that identifies the affiliate attacker in any action the user performs with the company's website.

Despite its recent widespread and importance, there are relative few quantitative works addressing affiliate marketing. A recent exception characterizes the relationship between social media presence of affiliates and their success in converting customers [Olbrich et al. 2019], indicating that affiliates with strong presence in online social networks have generate higher returns. In the direction of a quantitative exploration, this work contributes with the analysis of a real affiliate marketing program, providing insights about its network and revenue structure. Such characterization can later be used to drive incentive mechanisms and pricing for compensations.

7. Conclusion

Affiliate marketing is a fundamental marketing strategy for e-commerce with the potential to significantly boost sales and brand placement at low cost to a company. Not surprising, affiliate marketing is being deployed by an increasingly large number of companies under different rules and programs. Designing an effective affiliate marketing program requires determining the actions allowed by the affiliates and the compensation for such actions, such as selling a product and bringing new affiliates. Clearly, the design of an effective program requires understanding the dynamics of affiliates and their profiles.

The analysis of *Clube Hurb* presented her has focused on the affiliate network, the revenue generated by affiliates, and the relationship between network structure and

revenue. The findings are revealing since most affiliates have little to no participation in either constructing the network (90.7% are isolated nodes) or generating revenue (99.5% never sell). However, the small fraction that engages in the network activity or revenue activity move the marketing program. Moreover, the distribution of various characteristics among them is heavily skewed with an even smaller fraction of affiliates generating the vast majority of the revenue or bringing more affiliates to the program through referrals, as indicated by the Lorentz curves and Gini index measurements.

In some sense, the analysis has shown the *Clube Hurb* has two classes of affiliates: ones that join the program in the heat of the moment (the vast majority), and ones that join the program thinking about commissions (a small minority). Engaging the former class into making their first sale could significantly improve the revenue generated by the marketing program. Moreover, affiliates that join the network thorough referral have three times more chances to do at least one sale, and consequently generate much more revenue on average. While the reason for this significant increase remains unclear, a possibility is that affiliates are more likely to engage meaningfully with the program when invited by a trusted social contact, and that some social filter implicitly applies on the referral process.

Finally, this work provides a necessary step towards the design of more effective affiliate marketing programs. While it is yet not clear how *Clube Hurb* should be modified to be more effective (in terms of revenue or engagement), the data analysis provided in this work will drive the course of action. Moreover, results presented here may also feed theoretical models for affiliate marketing programs, leading to more realistic dynamics.

References

Chachra, N., Savage, S., and Voelker, G. M. (2015). Affiliate crookies: Characterizing affiliate marketing abuse. In *ACM Internet Measurement Conference (IMC)*, pages 41–47.

Daniele, R., Frew, A. J., Varini, K., and Magakian, A. (2009). Affiliate marketing in travel and tourism. In *Information and Communication Technologies in Tourism 2009*, pages 343–354.

Duffy, D. L. (2005). Affiliate marketing and its impact on e-commerce. *Journal of Consumer Marketing*, 22(3):161–163.

Dwivedi, Y. K., Rana, N. P., and Alryalat, M. A. A. (2017). Affiliate marketing: An overview and analysis of emerging literature. *The Marketing Review*, 17(1):33–50.

Goel, S., Watts, D. J., and Goldstein, D. G. (2012). The structure of online diffusion networks. In *ACM Conference on Electronic Commerce (EC)*, pages 623–638.

Gregori, N., Daniele, R., and Altinay, L. (2014). Affiliate marketing in tourism: determinants of consumer trust. *Journal of Travel Research*, 53(2):196–210.

Hinz, O., Skiera, B., Barrot, C., and Becker, J. U. (2011). Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6):55–71.

Ho, J. Y. and Dempsey, M. (2010). Viral marketing: Motivations to forward online content. *Journal of Business research*, 63(9-10):1000–1006.

Ivkovic, M. and Milanov, D. (2010). Affiliate internet marketing: Concept and application analysis. In *IEEE International Conference on Education and Management Technology*, pages 319–323.

Laudon, K. C. and Traver, C. G. (2019). *E-Commerce 2019: Business, Technology and Society*. Pearson, 15th edition.

Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1).

Nguyen, H. T., Thai, M. T., and Dinh, T. N. (2016). Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *ACM International Conference on Management of Data (SIGMOD)*, pages 695–710.

Olbrich, R., Schultz, C. D., and Bormann, P. M. (2019). The effect of social media and advertising activities on affiliate marketing. *International Journal of Internet Marketing and Advertising*, 13(1):47–72.

Reichstein, T. and Brusch, I. (2019). The decision-making process in viral marketing—a review and suggestions for further research. *Psychology & Marketing*, 36(11):1062–1081.

Snyder, P. and Kanich, C. (2015). No please, after you: Detecting fraud in affiliate marketing networks. In *ACM Workshop on the Economics of Information Security (WEIS*.

Suryanarayana, S. A., Sarne, D., and Kraus, S. (2019). Information disclosure and partner management in affiliate marketing. In *ACM International Conference on Distributed Artificial Intelligence (DAI)*, pages 1–8.

Watts, D. and Peretti, J. (2007). Viral marketing for the real world. *Harvard Business Review (HBR)*.