

Recommending Influencers to Merchants using Matching Game Algorithm

Gomes, J.M.¹ and Dias, L.A.V.¹

¹Instituto Tecnológico de Aeronáutica - ITA

December 20, 2023

Abstract

The goal of this work was to apply the “**Gale-Shapley**” algorithm to a real-world problem. We analyzed the pairing of influencers with merchants, and after a detailed specification of the variables involved, we conducted experiments to observe the validity of the approach.

We conducted an analysis of the problem of aligning the interests of merchants to have digital influencers promote their products and services. We propose applying the matching algorithm approach to address this issue.

We demonstrate that it is possible to apply the algorithm and still achieve corporate objectives by translating performance indicators into the desired ranking of influencers and product campaigns to be advertised by merchants.

Introduction

The “Digital Influencer” is a phenomenon that emerged around 2015¹ and represents an evolution of “bloggers” and “vloggers”. Based on the reach of these individuals, they have taken on roles as opinion leaders and have been enlisted by companies for merchandising projects [1].

Studies not directly related to recommendation systems address the clustering of consumers according to their preferences, defining profiles of consumption, attention, personal, or social objectives [2]. “**Gale-Shapley**” focused on the formation of stable pairs between any two groups, meaning the formed pairs are within the preference lists of each individual [3], and no pair is formed with individuals outside each person’s personal list.

Several variations of the algorithm have been created, such as:

1. Heterogeneous players: for example, men and women in marriage;

¹During the same period, the term “Content Creators” was also used

2. Homogeneous players: roommates;
3. $1 - 1$ relationships: again, marriage or roommates;
4. $m - 1$ relationships: residents and hospitals or students and projects [4].

In this work, we approach a recommendation system as a matching game where a finite number of influencers (those who recommend) and an also finite number of merchants and industrialists with their services and products need to be paired. Each influencer is willing to recommend items to their clients (in commerce and industry), who in turn have a limited supply of advertisements. Influencers, merchants, and industrialists are heterogeneous, but we can assume the recommendation itself as a homogeneous element. Each agent (influencer or consumer) derives utility from the attention received or given with additive objective functions. It is an $n - to - m$ two-sided model, with attention vectors, one for each recommendation, and an allocation of recommendations to listeners or readers, so that the demand of each is satisfied, and the number of recommendations made by the influencer does not exceed the available attention.

We applied the algorithm developed based on the work of D. Gale and L. S. Shapley [3] with the revisions by A. E. Roth [4, 5], using the implementation by H. Wilde *et al.* [6] as a reference. This was done in a proof of concept where we aimed to achieve the optimal matching between influencers, merchants, and their products and services.

Problem statement

Let $F = \{f_1, f_2, \dots, f_n\}$ be a set of influencers, $P = \{p_1, p_2, \dots, p_n\}$ a list of available products, and $V = \{v_1, v_2, \dots, v_3\}$ a list of merchants selling these products.

Each influencer f_i provides a preference list with items from P (see Table 1). If product p_j appears in the list of f_i , then f_i considers p_j desirable. We denote D_i as the set of products desired by f_i .

Id	f_i	Rep.	$r(f_i)$	p_1	p_2	\dots	p_j
296387a17f2b		148.0	437441	11534	\dots	-	
64c6cf346842		62.0	9976	572953	\dots	-	
89f3d85b7645		57.0	367959	621987	\dots	-	
89cca91ef90c		49.0	850416	0	\dots	-	
1a935a652ab0		18.0	673576	0	\dots	-	

Table 1: Sample of the list of influencers and their preferences

Each merchant v_k advertises a list of products P_k where P_1, P_2, \dots, P_l segments P (see Table 3). For each product announced in P_k , the merchant indicates how many influencers they want to allocate. Let $M_k = \{f_i \in F : P_k \cap D_i \neq \emptyset\}$ be the set M_k of influencers who wish to advertise the products offered by merchant v_k . Consider $r(f_i)$ as the reputation of the influencer. Thus, for all $p_j \in P_k : \nu$, we represent ν_k^j as the preferences of v_k for product p_j - obtained by

Id	f_i	Rep.	$r(f_i)$	p_1	p_2	\dots	p_j
296387a17f2b			79.0	437441	11534	\dots	-
64c6cf346842			44.0	9976	572953	\dots	-
89f3d485b7645			26.0	367959	621987	\dots	-
89cca91ef90c			14.0	850416	0	\dots	-
1a935a652ab0			8.0	673576	0	\dots	-

Table 2: Sample of the list of influencers and their preferences classified by Average Purchases

excluding all influencers f_i not interested in p_j in the order dictated by $r(f_i)$. At the same time, while each merchant v_k is limited to a quantity q_k of influencers they can accept (see Table 3), each product p_j also has a limit l_j on how many influencers it can be offered to.

Id	v_k	Quota	q_k
43d8ac5e2f82			45
c8923e87aa1f			3
b7b3db8ade87			2
7f370a5d8046			1
6c6d30a69b40			1

Table 3: Sample of the list of merchants

Cod	p_j	Quota	l_j	Com.	v_k
823128			43	43d8ac5e2f82	
12586			16	43d8ac5e2f82	
437441			16	43d8ac5e2f82	
7976			15	43d8ac5e2f82	
17150			12	43d8ac5e2f82	

Table 4: Sample of the list of products by merchant

In Tables 1, 3 and 4, we present a sample where we can observe that each influencer can represent one or more products, and each merchant can accommodate up to a certain number of influencers and offer their products, which in turn can accommodate a specific number of influencers simultaneously. The total number of desired influencers may be less than, equal to, or greater than the supported number, but we stipulate $\max\{l_j : p_j \in P_k\} \leq q_k \leq \sum\{l_j : p_j \in P_k\}$ as the operational limit.

An allocation E is a subset of $I \times P$ such that $(f_i, p_j) \in E \implies p_j \in D_i$, meaning that f_i wants to advertise p_j , and for each influencer $f_i \in F$, $|\{(f_i, p_j) \in E : p_j \in P\}| \leq l_j$. Influencer f_i has been recommended product p_j , and product p_j has been recommended to influencer f_i if $(f_i, p_j) \in E$. We can also say that if f_i has been recommended p_j in E , where $p_j \in P_k$, then f_i has been recommended to v_k , and v_k has been recommended to f_i .

If an influencer $f_i \in F$, f_i has been paired in E with some product p_j , then $E(f_i)$ represents p_j , or conversely, f_i does not exist in E .

For a given product, the expression $p_j \in P$, $E(p_j)$ represents the set of influencers allocated with p_j in E . Product p_j can be under-allocated, fully allocated, or over-allocated if $|E(p_j)|$ is respectively less than, equal to, or greater than l_j . Similarly, for any merchant $v_k \in V$, $|E(v_k)|$ represents the set of influencers recommended to v_k in E . Merchant v_k is under-allocated, fully allocated, or over-allocated if $|E(v_k)|$ is respectively less than, equal to, or greater than q_k .

A matching E is an allocation such that the allocation of each product satisfies the condition $p_j \in P, |E(p_j)| \leq l_j$, and the allocation of each merchant satisfies the condition $v_k \in V, |E(v_k)| \leq q_k$.

Thus, p_j is allocated to a maximum of l_j influencers in E as long as it does not exceed q_k for v_k in E .

The pair $(f_i, p_j) \in (F \times P)$ is subject to $p_j \in D_i$ (influencer f_i desires p_j) and $f_i \notin E$ or f_i prefers p_j over $E(f_i)$.

Both p_j and v_k are under-allocated, or p_j is under-allocated and v_k is fully allocated, and either $f_i \in E(v_k)$ or v_k prefers f_i to the least qualified influencer in $E(v_k)$, or p_j is fully allocated and v_k prefers f_i to the least qualified influencer in $E(p_j)$ - we say this case of (f_i, p_j) is a blocking pair in E . A matching is stable if there are no blocking pairs.

Proof of Concept

Approach Description

To extract lists of potential influencers, products, and merchants available in the dataset, performance indicators were simulated as a way to demonstrate an application of the solution:

1. Influencers f_i , reputation $r(f_i)$, and list of desired products E based on different prioritization criteria:
 - (a) Reputation $r(f_i)$ from the *GT (Total Expenditure)*, i.e., the more the consumer spends on a particular product, the higher their reputation for that product (see Table 1)
 - (b) Reputation $r(f_i)$ from the *FMC (Frequency of Average Purchase)* (see Table 2), given by the formula:

$$r(f_i) = K/U,$$

where:

- i. K : Number of orders
- ii. U : Total consumers or potential influencers

2. Merchants v_k and their respective quotas q_k :

- (a) “Quota” q_k calculated based on the total volume sold, i.e., the more the merchant v_k sells, the more space for influencers they will have (see Table 3)

3. Products p_j , quota l_j per product and merchant v_k :
- (a) Quota l_j for product p_j assigned manually, i.e., based on the total quota l_j of merchant v_k , they can request influencers for their products (see Table 4)

Note that the value assigned to $r(f_i)$ is not important - the descending order of this value is sufficient for the algorithm to prioritize an influencer.

Experiments

To validate this proposal, we collected data from different sources:

1. Public datasets from the internet²
- (a) Electronics sales[7]
- (b) Supermarket sales[8]

The importance of sorting the data lies in the fact that, in general, the "Gale-Shapley" algorithm terminates in most cases in n^2 iterations, with a worst-case computational time of $\Omega(n^2)$ [9]. Thus, it is desirable to reduce the number of elements analyzed to consider only the most interesting ones. For this purpose, we use criteria based on performance indicators (see Approach Description) aligned with organizational objectives.

In the supermarket database, there is no identification of the buyer (i.e., potential influencer). In this situation, a profile was defined composed of the consumer's information such as *branch*, *city*, *type*, and *gender*.

Results

Kaggle Dataset - Electronics Store - Average Purchase Frequency X Total Spending

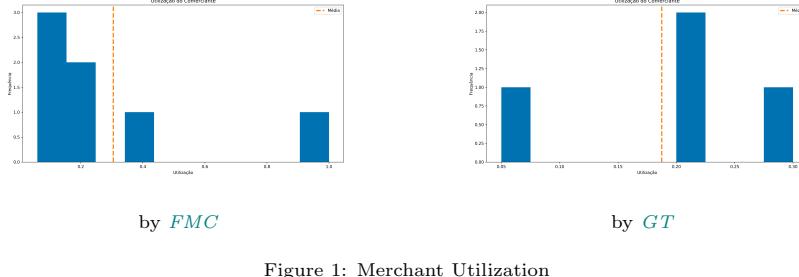
Using data from the electronics store[7], we observe in Figure 1 the low efficiency in recommending influencers to merchants. This result is a consequence of the fact that electronic products are not repeatedly purchased by the same consumer.

Given the sales volume of some manufacturers (see Table 3), the maximum quota of influencers ends up being high due to $\sum\{l_j : p_j \in P_k\}$ (see Problem statement). This is reflected in the high frequency of empty spaces that we can observe in Figure 2.

In Figure 3, we see a concentration in the low utilization of advertisements. The wide variety of different products does not allow for specific concentration.

With the assignment of a preference for consumers and as a consequence of the great diversity of different products, it is noted that the assignment is not able to indicate consumer preference (as can be seen in Figure 4).

²See <https://www.kaggle.com>



by *FMC*

by *GT*

Figure 1: Merchant Utilization

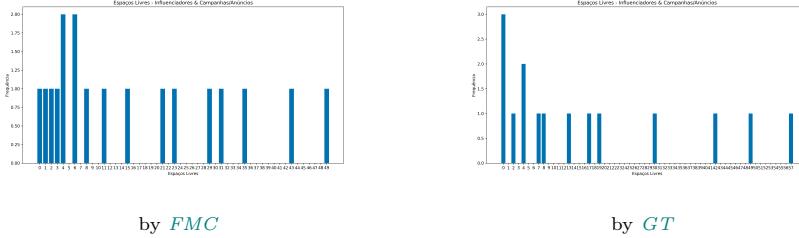


Figure 2: Free slots

The consumer classification by *FMC* or *GT* is unable to bring a significant change in the assignment of their preference in indicating them as a potential influencer for a particular product (see in Figure 5).

Kaggle Dataset - Supermarket - Average Purchase Frequency vs Total Spending

Using supermarket sales data[8], we can see in Figure 6 the separation between unpopular products on the left with low ad recommendation frequency and popular products on the right.

The initial recommendations for the consumer are mostly satisfied, with some extreme cases of many empty spaces, as observed in Figure 7.

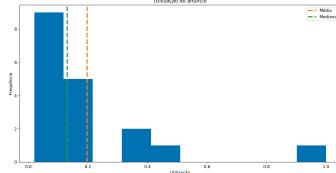
In Figure 8, we observe the distribution on the left of unpopular products and on the right of popular products in the recommendations for consumers.

We can see in Figure 9 that there was no clear indication of consumer preference.

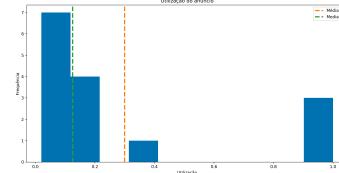
The consumer classification does not change the fact that it was not possible to pair the consumer with a specific advertisement (see Figure 10).

Conclusions

The “**Gale-Shapley**” algorithm is a mature solution, not only effective in forming optimal pairs but also efficient in its applicability to a real-world problem.

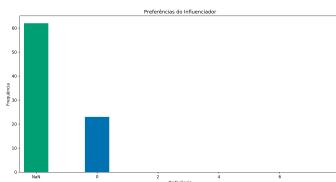


by *FMC*

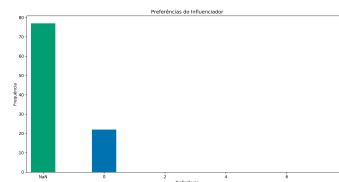


by *GT*

Figure 3: Advertisement Usage



por *FMC*



por *GT*

Figure 4: Preferência do influenciador

Generic conclusions

In a general context, this work demonstrates a practical application of the “Gale-Shapley” algorithm, and the solution found meets a real need.

0.1 Generic conclusions

In particular, this work solves a problem of pairing online media influencers with product and service campaigns from merchants interested in improving their performance in current media.

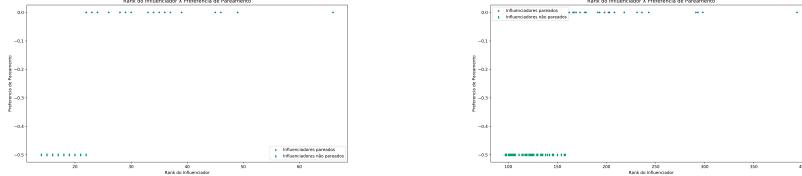
Acronyms

FMC Frequency of Average Purchase - pages: 4, 6–10

GT Total Expenditure - pages: 4, 6–10

References

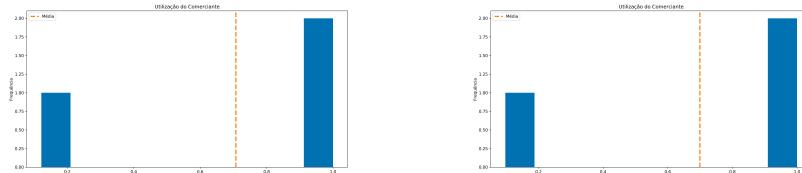
- [1] I. Karhawi *et al.*, “Influenciadores digitais: Conceitos e práticas em discussão,” *Communicare*, vol. 17, no. 12, pp. 46–6, 2017.



por *FMC*

por *GT*

Figure 5: Rank influenciador e preferência



by *FMC*

by *GT*

Figure 6: Merchant Utilization

- [2] A. Kanavos, S. Iakovou, S. Sioutas, and V. Tampakas, “Large Scale Product Recommendation of Supermarket Ware Based on Customer Behaviour Analysis,” *Big data and cognitive computing*, vol. 2, no. 2, pp. 11–, 2018.
- [3] D. Gale and L. S. Shapley, “College admissions and the stability of marriage,” *The American Mathematical Monthly*, vol. 120, no. 5, pp. 386–391, 2013.
- [4] A. E. Roth, “The college admissions problem is not equivalent to the marriage problem,” *Journal of economic Theory*, vol. 36, no. 2, pp. 277–288, 1985.
- [5] A. E. Roth and M. Sotomayor, “Two-sided matching,” *Handbook of game theory with economic applications*, vol. 1, pp. 485–541, 1992.
- [6] H. Wilde, V. Knight, and J. Gillard, “Matching: A python library for solving matching games,” *Journal of Open Source Software*, vol. 5, no. 48, p. 2169, 2020.
- [7] “Ecommerce purchase history from electronics store.” (2020), [Online]. Available: <https://www.kaggle.com/datasets/mkechinov/ecommerce-purchase-history-from-electronics-store>.
- [8] “Historical record of sales data in 3 different supermarkets.” (2019), [Online]. Available: <https://www.kaggle.com/datasets/aungpyaearp/supermarket-sales>.
- [9] J. Kleinberg and E. Tardos, *Algorithm design*. Pearson Education, 2006.

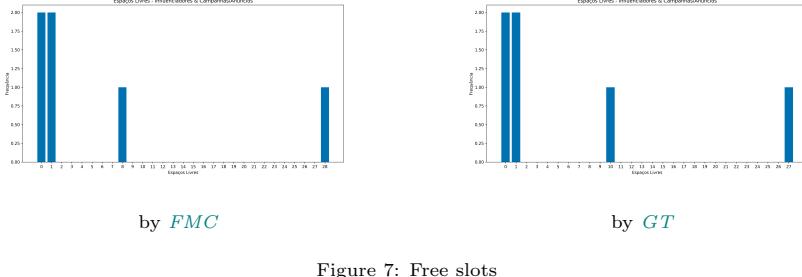


Figure 7: Free slots

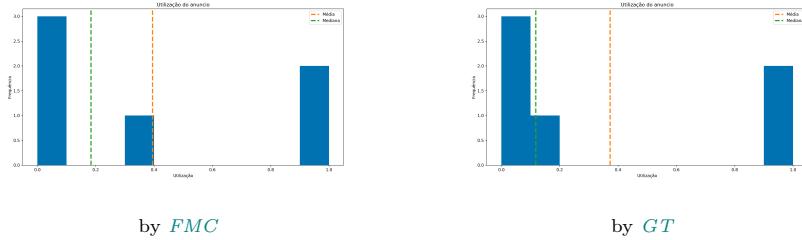


Figure 8: Ad Utilization

- [10] P. Szabó and B. Genge, “Efficient conversion prediction in e-commerce applications with unsupervised learning,” in *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2020, pp. 1–6. doi: 10.23919/SoftCOM50211.2020.9238344.
- [11] Ecossistema Ltda., “Wibx whitepaper,” 2018.

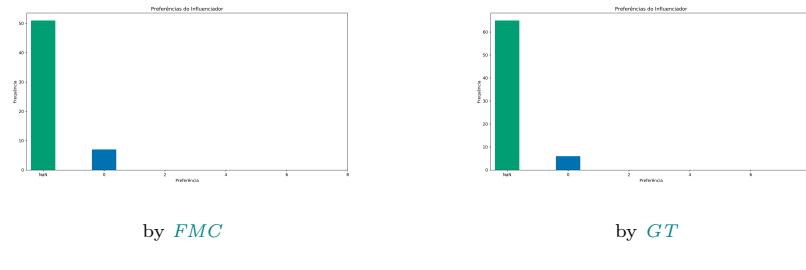


Figure 9: Influencer Preference

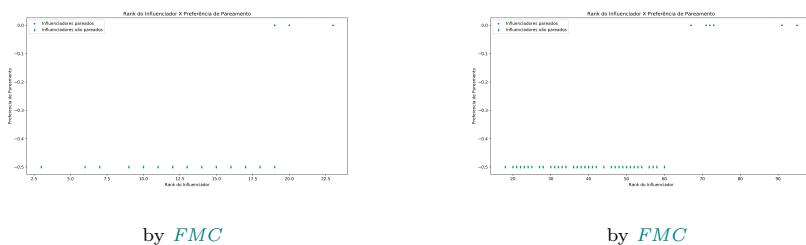


Figure 10: Influencer Rank and Preference