Physica A 524 (2019) 36–44

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Market basket analysis by solving the inverse Ising problem: Discovering pairwise interaction strengths among products

Mauricio A. Valle [a,*], Gonzalo A. Ruz [b], Sergio Rica [b]

[a] Facultad de Economía y Negocios, Universidad Finis Terrae, Santiago, Chile
[b] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Santiago, Chile

HIGHLIGHTS

- Ising model was used to analyze the purchase behavior.
- Collective purchasing system has a similar behavior of a spin glass.
- Second order interactions explain on average 78% of the purchase information.
- Hierarchical structures are recovered using the minimum spanning tree from the couplings.

ARTICLE INFO

Article history:
Received 29 December 2018
Received in revised form 22 February 2019
Available online 12 March 2019

Keywords:
Inverse Ising problem
Boltzmann machine
Transactional data base
Pairwise interaction
Minimum spanning tree
Purchase pattern

ABSTRACT

Large datasets containing the purchasing information of thousands of consumers are difficult to analyze because the possible number of different combinations of products is huge. Thus, market baskets analysis to obtain useful information and find interesting pattern of buying behavior could be a daunting task. Based on the maximum entropy principle, we build a probabilistic model that explains the probability of occurrence of market baskets which is equivalent to Ising models. This type of model allows us to understand and to explore the functional interactions among products that make up the market offer. Additionally, the parameters of the model inferred using Boltzmann learning, allow us to suggest that the buying behavior is very similar to the spin-glass physical system. Moreover, we show that the resulting parameters of the model could be useful to describe the hierarchical structure of the system which leads to interesting information about the different market baskets.

1. Introduction

Since the last twenty years, statistical physics models have been used to understand and discover properties of economic systems. For example, some pioneering studies have been able to detect hierarchical structures between different stocks traded in financial markets, using tools and procedures developed to model physical systems by determining correlations among assets [1–3]. Some of these studies have had a greater impact on quantitative finance, introducing new methodologies and models to explain price fluctuations [4,5], to identify the noise in financial correlation matrices to improve portfolio optimization [6], also to describe the dynamics and properties of assets in which scaling laws have been found [7–9].

* Corresponding author.
E-mail address: mvalle@uft.cl (M.A. Valle).

https://doi.org/10.1016/j.physa.2019.03.001
0378-4371/

In the field of marketing and consumer behavior, there are large volumes of data present in transactional databases that record each purchase of a customer over time. The availability of this data makes it interesting to study purchasing patterns at an aggregate level. A recent area of research on purchasing patterns using networks has revealed that the buying phenomenon exhibits scale properties. The networks of products built from the list of transactions tend to possess the typical scale-free network [10–12]. However, unlike quantitative finance, there appears to be no further development in the analysis of purchasing behavior using more sophisticated techniques.

In this paper, we investigate purchases from transactional databases. These data contain the information of market preferences that can be seen as a complex system composed of many spins (or market offers) that interact with each other. The products purchased, reveal the preferences of the consumers of a set of various brands and products available at a given time. In a physical sense, a market basket is equivalent to a spin configuration in the system. Our approach is based on modeling purchases statistically using the maximum entropy principle [13,14]. The resulting model allows us to understand how purchasing units interact with each other and to observe recurring behavior patterns in different product samples.

The maximum entropy principle has been a useful tool for studying scale and structure properties in neurosciences. For instance, it is possible to model collective neurons activations under different stimuli [15–18]. In these cases, the idea is to characterize the distribution of the physical system by taking the first moments and the pairwise correlations between the variables of the observed distribution, and use them as constraints. Then through a process of entropy maximization, it is desirable to find a distribution that is consistent with those moments and correlations, but does not assume the existence of higher order interactions. This kind of distribution follows the Boltzmann distribution. The inference of the parameters of this distribution is difficult in general, and it corresponds to the inverse Ising problem [19].

In this line, we demonstrate that the Ising model approach is also feasible for describing the aggregate purchasing behavior of a large number of consumers. This approach allows an economic interpretation, in which the inferred parameters of the distribution can be understood as a measure of economic activity among the elements that constitute the system [20].

The results indicate that this kind of models are able to recover satisfactorily the distribution of the states. This makes it possible to validate the hypothesis that the pairwise interactions contain enough information to explain the phenomenon of purchases, as has been discovered in financial markets and in the functioning of biological neurons. As an example of the application of the inferred parameters, we use the couplings to obtain the hierarchical structure of the collective purchasing system.

2. Model and learning

Using the empirical data, it is necessary to find a probability distribution whose number of parameters is less than the possible number of purchase states. The states of the collective system of purchases are all possible combinations of elements that make up a market basket. Under the approach of the Ising-based model, we represent a market basket with the state $\mathbf{s} = (s_1, \ldots, s_N)$. Each spin $s_i = \pm 1$ represents the presence or absence of the product i in the market basket. The probability of observing one of these states is described by:

$$P(\mathbf{s}) = \frac{1}{Z} \exp\left(-\beta H(\mathbf{s})\right). \tag{1}$$

We denote $P(\mathbf{s})$ as the Gibbs distribution. Here $H(\mathbf{s})$ represents the Ising energy of the system in which we consider only the pairwise interactions and not the higher orders. The parameter β (the inverse of the temperature) represents the fluctuations or randomness. For large temperatures ($\beta \to 0$), thermal fluctuations dominate the system and all spin configurations have equal probability, whereas long range order and aligned spins domains are observed for smaller temperature than a critical temperature T_c ($\beta > \beta_c$). Finally, in the zero temperature limit ($\beta \to \infty$) spins take a well defined value. This fact will become more precisely later-on, according to the transition probability (7). The constant Z is the partition function that ensures normalization, and $H(\mathbf{s})$ is:

$$H(\mathbf{s}) = -\sum_i h_i s_i - \frac{1}{2} \sum_{i<k} J_{ik} s_i s_k, \tag{2}$$

which represents the Hamiltonian or the energy of the system in the configuration $\mathbf{s}$. The interaction between a product i and another k is represented by the couplings J_{ik}. These parameters represent a measure of influence of one product on another. If $J_{ik} > 0$, both products tend to be present simultaneously in the market basket, while $J_{ik} < 0$, one product is present while the other is not (equivalent to say that the spins are in opposite directions). The external field (the magnet field in the usual Ising model) of ith product is h_i. It can be considered as the effect of external agents to the system that influences the product i [20]. The Gibbs probability depends explicitly on the products of the parameters h_i, J_{ik} and β, therefore we cannot determine all parameters but the products: $h_i^* = \beta h_i$ and $J_{ik}^* = \beta J_{ik}$ which are the ones that have relevance. From now on we drop the $*$ in these parameters. Some configurations are more likely to occur than others, not only because of every single spin (product), but also because of the interrelationship of those products with others in the same market basket. The probability of occurrence of the configurations is directly related to the energy of the system expressed in Eq. (2). Thus, the J_{ik} couplings of the system have consequence in the realization of different market baskets. The process of inferring parameters h_i^* y J_{ik}^* from data is known as the inverse Ising problem [19]. The result of the inference gives an approximation to the pairwise distribution $P(\mathbf{s})$ in which:

$$\langle s_i \rangle_{Ising} = \langle s_i \rangle_{data} \quad \forall i, \tag{3}$$

$$\langle s_i s_k \rangle_{Ising} = \langle s_i s_k \rangle_{data} \quad \forall i \neq k. \tag{4}$$

Our approach used to learn the distribution, follows the principle of maximum entropy [13,14] which has various applications such as neuroscience and neuro-computing, particularly in cases in complex systems where there are a limited amount of data available, and it is necessary to make inferences on the states of the total system with these data [21]. We consider a Boltzmann learning process [22] in which the parameters of $H(\mathbf{s})$, the fields and couplings, are inferred iteratively through successive corrections calculated as:

$$h_i^{\tau+1} = h_i^{\tau} + \nu\left(\langle s_i \rangle_{data} - \langle s_i \rangle_{Ising}^{\tau}\right), \tag{5}$$

$$J_{ik}^{\tau+1} = J_{ik}^{\tau} + \nu\left(\langle s_i s_k \rangle_{data} - \langle s_i s_k \rangle_{Ising}^{\tau}\right). \tag{6}$$

where ν is the learning rate parameter, and it is understood that the Ising averages are done with the parameters h_i^{τ} and J_{ik}^{τ}. We use as the initial values for the field and the couplings, the mean of the activation of the spins and the correlation between spins, respectively, i.e., $h_i^{\tau=0} = \langle s_i \rangle_{data}$ and $J_{ik}^{\tau=0} = \rho_{ik}$ being ρ_{ik} equal to $\left(\langle s_i s_k \rangle - \langle s_i \rangle \langle s_k \rangle\right) / \left((\langle s_i^2 \rangle - \langle s_i^2 \rangle)(\langle s_k^2 \rangle - \langle s_k^2 \rangle)\right)^{1/2}$ (see Section 5 for more details).

The transition probabilities of the state changes were simulated using the Metropolis–Hasting (MH) algorithm. Thus, spin states swap $\mathbf{s} \to \mathbf{s}'$, are done by comparing with a random number generated uniformly in the interval [0, 1] [23], being the transition probability given by:

$$\Pi(\mathbf{s} \to \mathbf{s}') = \begin{cases} 1 & \text{if} \quad \Delta E \leq 0 \\ \exp(-\beta \Delta E) & \text{otherwise} \end{cases}, \tag{7}$$

where $\Delta E = H(\mathbf{s}) - H(\mathbf{s}')$ is the energy difference between the initial state $\mathbf{s}$ and the final state $\mathbf{s}'$. In this way, it is possible to compute the observables $\langle * \rangle_{Ising}$. This process is iterated as many times as necessary until a certain level of accuracy is reached, or up to a predetermined number of iterations of the Boltzmann learning.

3. System of collective purchases

3.1. The data

To carry out our analysis, we took a transactional database of 179 610 purchase records from 17 093 regular customers of a branch of a supermarket chain in Santiago, Chile. These data involve purchases between July 2010 and November 2011. This branch is located in a sector of high public affluence, and it is usually used by consumers to make on-the-spot purchases, i.e., purchases of a low number of products and not for the end-of-the-month shopping to replenish household food.

The database contains detailed purchasing information of the available products. We have considered as a unit of analysis, the subcategories of products, which in total are 220. However, we took the 21 top sellers subcategories. For our purposes, these subcategories are equivalent to the individual elements of the collective system of purchases.

To analyze the results of the inverse Ising problem, we took $M = 250$ samples from the transactional dataset, each one with a different combination of subcategories, but keeping always $N = 12$ of them. In this way, it is possible to have different number of samples to compare results under different combinations of purchases and products. The range of the number of transactions in the samples varies between 23 487 and 70 921.

3.2. Small correlations for purchases

To get an idea of the means and correlations between subcategories of products, we have taken a sample M_i of the set of the 250. For this sample we estimated the density of the means of the product occurrences of each product, and their pairwise correlations. In Fig. 1**A** it is observed that in general, subcategories tend to be present in less than 30% of the total of the market baskets, while few products tend to have greater presence. Fig. 1**B** illustrates the density of the correlations between products. It is interesting to note that the correlations are small. 95% of the correlations are between −0.19 and 0.18, with a mean of −0.035. In fact, these correlation magnitudes tend to be small in the same way for all other samples. The 95% interval of the mean correlations of all samples is [−0.053, −0.015].

These results suggest that the presence or activation of a subcategory (i.e., the purchase of a product belonging to that subcategory), is independent of what happens to other products in other subcategories. This would indicate that the system that governs the phenomenon of collective purchasing, is independent of the elements that constitute it, or that there are no correlations between the products. If we assume a model in which the probability of activation of a product is independent of another product, then the probability $p(K)$ that K products are active in a market basket[1] would

[1] The total number of active products in a basket is $K = \frac{1}{2}\left(N + \sum_{i=1}^{N} s_i\right)$.

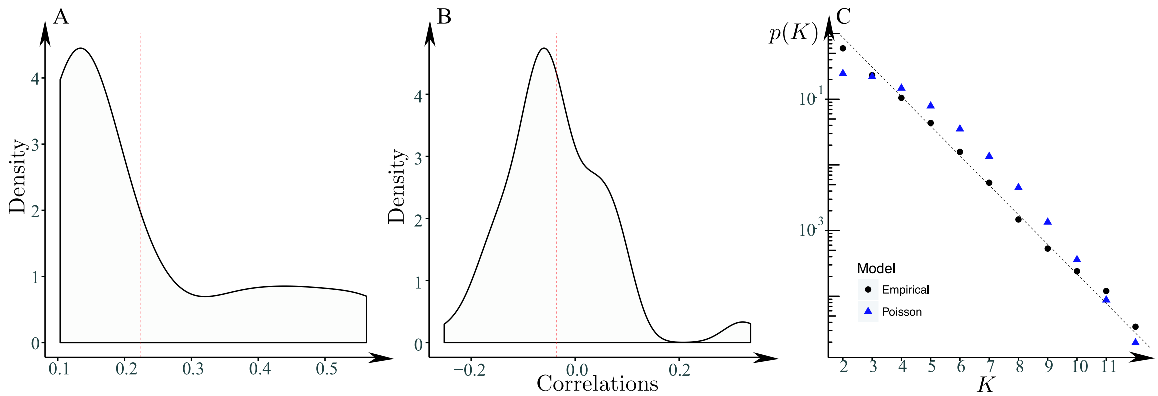

Fig. 1. **A**, Density of means of product occurrences of subcategory products for one of the M transactional sets, $\langle s_i \rangle = 0.223$. **B**, Density of correlations between pair of subcategory products of the sample, $\langle C_{ij} \rangle = -0.035$. The dashed red lines indicates the mean of the distributions. **C**, Probability distribution of the number of subcategory activations K of the empirical model (black) and of the independent model (blue) computed assuming a Poisson distribution. The last one was computed taking individual activation probabilities for each subcategory of the sample. The y-axes is in logarithmic scale.

approximate a Poisson distribution. This distribution is represented in Fig. 1**C** by blue triangles, while the distribution with black dots is from the empirical data of a selected sample. The distribution $p(K)$ of the empirical data approximates an exponential distribution: $p(K) \sim e^{-\lambda K}$ with the parameter $\lambda = 2.26 \pm 0.08$ as a result of the fit. On the other hand, the Poisson distribution that assumes independence between spins, tends to overestimate the probability in a mid range of K, while for more extreme values, the distribution tends to underestimate these probabilities. These differences suggest that the market baskets do not appear to be a result of a behavior that assumes complete independence between its elements.

This makes us suspect that even when correlations between pairs of subcategories are small, there are higher order interactions or multicorrelation between products that cannot be captured with simple linear correlations. In this way, it is possible that the correlations between different product groups, reinforce each other as a system of interconnected network. In this way, we opted for a maximum entropy model that brings us closer to the distribution of the 2^N possible states that make up the empirical market baskets.

4. Model consistency

To show that the pairwise model of maximum entropy $P(\mathbf{s})$ is consistent with the actual transactions, we carry out, for each of the 250 samples, a Boltzmann learning process of 2000 steps with an initial learning rate $\nu = 0.8$ with a decay of 0.01. In each of these Boltzmann learning steps, we computed the means and correlations from 40 000 Monte Carlo samplings steps, using binomial initial configuration as $s_i = \pm 1$ with probabilities 1/2, 1/2.

We track the mean squared error (MSE[2]) between the data and the reconstructions $\langle s_i s_k \rangle_{data} - \langle s_i s_k \rangle^{\tau}_{Ising}$ and $\langle s_i \rangle_{data} - \langle s_i \rangle^{\tau}_{Ising}$. While the number of Boltzmann machine iterations is fixed, we have ensured that with this number of steps, the MSE is always less than 10^{-3}.

4.1. Means and correlations recovery

To test consistency between the model learned and the empirical data, for each sample, we computed the means $\langle s_i \rangle$ and the two-body connections $\langle s_i s_k \rangle$ of the empirical sample, and one from a simulated sample ($\langle s_i^* \rangle$ and $\langle s_i^* s_k^* \rangle$) using the learned $P(\mathbf{s})$. The results are illustrated in Fig. 2 for a specific sample M_i.

Fig. 2**A** and **B** show the plot between the means and the recovered pairwise connections for a randomly chosen sample of transactional data M_i. In both cases the root mean square error (RMSE) was 0.0026 and 0.0052 respectively with correlations above 0.9. For the correlations C_{ik}, Fig. 2**C**, shows that the situation is similar. The RMSE was 0.0075 and correlation was 0.801. The comparison between the observed components and those recovered from the distribution of maximum entropy, suggests that the model was capable to recover successfully the orientations and covariances of the collective purchasing system.

A general picture of the recovery of the pairwise connections is illustrated in Fig. 3**A** and **B**, in which we included the results of all sets of 250 samples. As each sample M_i contains a different set of $N = 12$ subcategories of products, we have different measurements of pairwise connections C_{ik} for the same pair of products i and k. These values are averaged and

[2] For example, for pair correlations, $MSE = \sum \left(\langle s_i s_j \rangle_{data} - \langle s_i s_j \rangle^{\tau}_{Ising} \right)^2 / T$ where T is the total number of pair correlations. The summation is on all pairs of combinations between spin i and j.

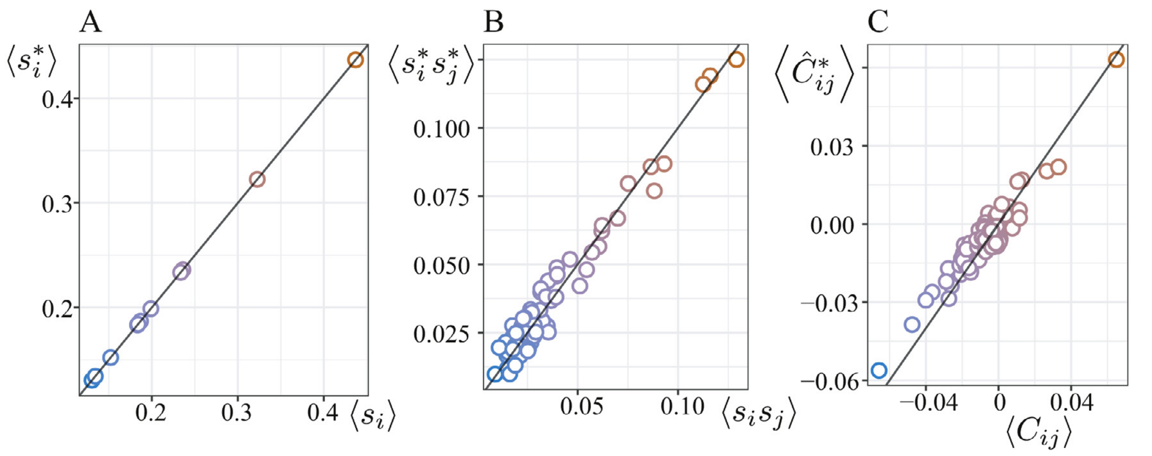

Fig. 2. **A**, Recovered means $\langle s_i \rangle$ from a MH sampling versus empirical ones. Straight line indicates equality. RMSE = .0026, ρ =.999; **B**, Recovered two-body connections $\langle s_i s_k \rangle$ from a MH sampling versus empirical ones. RMSE = 0.0052, $\rho = 0.979$; **C**, Recovered correlations $C_{ik} = \langle s_i s_k \rangle - \langle s_i \rangle \langle s_k \rangle$ from a MH sampling versus empirical ones. RMSE = 0.0075, $\rho = 0.801$.

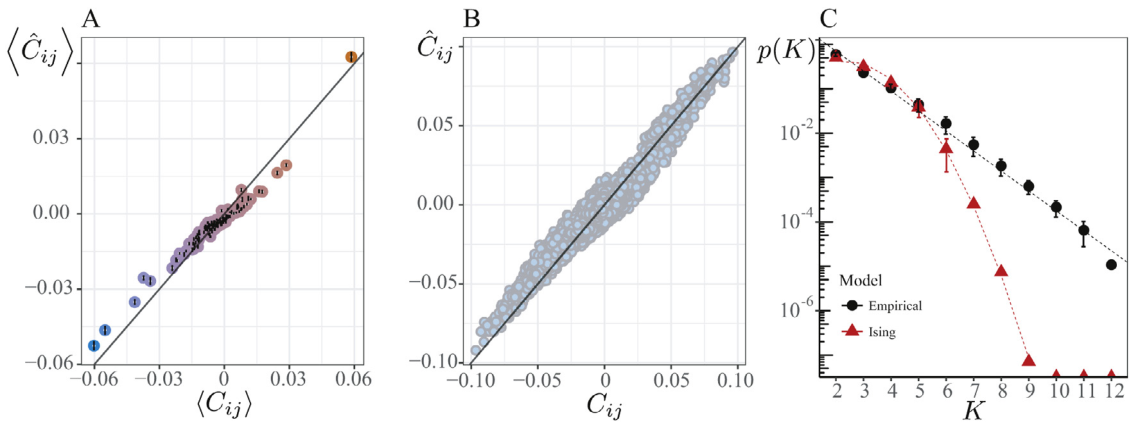

Fig. 3. **A**, Mean of the recovered two-body connections versus empirical ones from 250 different MH simulations of each one of the samples, RMSE = 0.0028, $\rho = 0.976$; **B**, Pooled version of **A**, all correlations were plotted together; **C**, Distribution of the number K of simultaneous product activations in the market baskets. Error bars represent the standard deviations over the set of $M = 250$ samples with $N = 12$ subcategories. Black line represents the empirical data, and the red points were obtained from Metropolis–Hasting simulations for every sample.

a standard deviation is calculated, which is plotted in Fig. 3**A**. In Fig. 3**B**, we did not take the average and we just plotted all the values of C_{ik} independently. These two graphs allow us to have an appreciation of the capability of the pairwise model $P(\mathbf{s})$ to retrieve these connections.

Fig. 3**C**, shows in black color the distribution of the number of activated subcategories K in the market baskets for all the set of the $M = 250$ samples. This behavior approximated to an exponential distribution $p(K) \sim e^{-\lambda K}$ with $\lambda = 2.44 \pm 0.03$. We also include the distribution from the inferred pairwise model (in red color). The dashed red line represents a Gaussian fit, $p(K) \sim e^{-\alpha K^2 + \beta K}$ with $\alpha = 4.53 \pm 0.19$ and $\beta = -14.26 \pm 0.19$. The pairwise model predicts a good distribution of simultaneous activations of products. However, from $K \geq 6$ it is possible to observe a complete divergence between the two distributions, which makes us think that influence of higher-order effects, are determining the behavior for market baskets with the higher number of products. Thus, the pairwise model $P(\mathbf{s})$ (Eq. (1)) seems to be a good model to explain the collective phenomenon of shopping but for market baskets of size $K \leq 5$ products.

4.2. Inferred fields and couplings

The relationship between h_i and s_i, and the relationship between C_{ik} and J_{ik} are not trivial as can be seen in Fig. 4. Taking into account all the parameters of the 250 samples, 95% of the field values are between −1.33 and 0.80, with a mean of $\langle h_i \rangle = -0.589$, while for the couplings these values correspond to −0.98 and 0.25 with a mean of $\langle J_{ik} \rangle = -0.422$. This result indicates that $\langle h_i \rangle < 0$ and, more important, the mean of coupling constant is also negative $\langle J_{ik} \rangle < 0$. Because

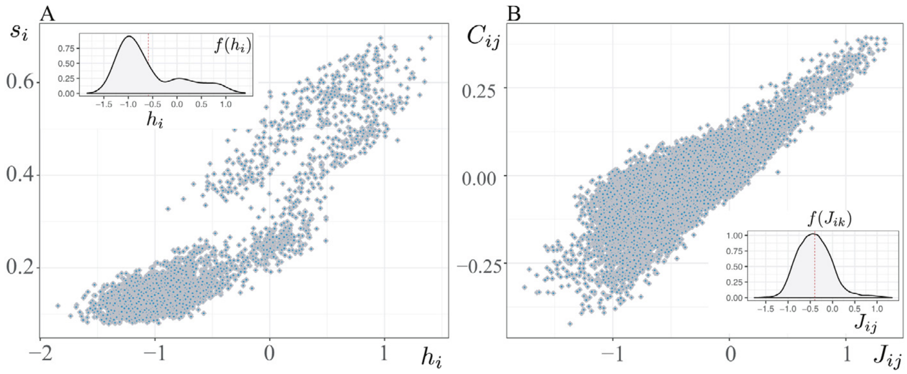

Fig. 4. **A**, Inferred field h_i for all set of samples M against their respective means orientations s_i; **B**, Inferred couplings J_{ik} for all set of samples M against their respective pairwise connections C_{ik}.

of the distributed nature of the coupling constant there are similarities with the spin-glass behavior. A spins-glass is a disordered system in which two spins can interact in a ferromagnetic way (the interaction allows a decrease of the energy by the alignment of the spins), or in an antiferromagnetic way (the interaction allows an increase of the energy by the anti-alignment of the spins). The distribution of the couplings $f(J_{ik})$ is very similar to a Gaussian, being the value of the negentropy[3] equal to $J(f(J_{ik})) = 0.0138 \pm 0.00031$. As these couplings are quenched, this results in some ground states of the system that could be frustrated, i.e., there will be indefinite spins orientations due to the random nature of the distribution of the couplings, where some lead to ferromagnetic and antiferromagnetic interactions at the same time over a spin. This results in low-energy states with energies very close to the global minima of the system [25].

Fig. 4**A** shows that there is a positive relationship, as expected, between the orientations and the fields h_i. Something similar happens in Fig. 4**B** for the relationship between pairwise connections and couplings J_{ik}. These results show that the fields and couplings portray much more information than only the linear dependence between the products of the collective purchasing system.

4.3. Are data explained by pairwise interactions?

The maximum entropy distributions inferred with the Boltzmann learning process seem to be satisfactory in recovering the net orientations and pairwise correlations between the subcategories of products. However, this tells us nothing about the influence of higher order interactions between the variables. To evaluate the ability of the pairwise model $P_2 = P(\mathbf{s})$ (In this section we added explicitly a sub-index 2 to the probability (1) to underline that it refers to a pairwise probability.) to capture intrinsic information in the collective purchasing system, we computed the Kullback–Leibler divergence [26] between the empirical distribution $P_{emp}(\mathbf{s})$ and the maximum entropy distribution $P_2(\mathbf{s})$, $D_{KL}(P_{emp} \parallel P_2)$. Then, we compared this value with the Kullback–Leibler divergence between the empirical distribution and the distribution of the independent model, i.e., Poisson distribution[4] P_1, $D_{KL}(P_{emp} \parallel P_1)$, with:

$$D_{KL}(P_{emp} \parallel P_2) = \sum_{i=1}^{2^N} P_{emp}(\mathbf{s}_i) \log \frac{P_{emp}(\mathbf{s}_i)}{P_2(\mathbf{s}_i)} \tag{8}$$

where $\mathbf{s}_i$ is one of the 2^N states of the system. The Kullback–Leibler divergence is also equivalent to the entropy (S) difference between the interacting variables of the system, and the variables taken independently, which has been called multi-information $I_N = S(P_1) - S(P_{emp})$ [27]. This indicates the total amount of information due to interactions, while $I_2 = D_{KL}(P_1 \parallel P_2) = S(P_1) - S(P_2)$ measures the amount of information due to second-order interactions. The ratio I_2/I_N gives us an indication of how much information is explained by second-order interactions.

The mean of I_2/I_N for all 250 samples was 0.779, while 95% of these measures are between 0.483 and 0.985. The distribution of the ratios can be seen in Fig. 5**A**, top right. Although that result indicates that the pairwise distribution explains most of the available information, we also see that 3.2% of the ratios are below 50% which would indicate that in these cases, the pairwise correlations are not enough to explain the purchasing information at an aggregate level. There is information that the maximum entropy model does not capture, and it is primarily due to higher order effects, particularly

[3] Negentropy is a distance measure from normality defined as $J(y) = H(y_g) - H(y)$ where $H(y)$ is the entropy of the data density, and $H(y_g)$ is the equivalent entropy of a Gaussian density with the same mean and variance of y. The value of the entropies were computed using a plug-in estimator [24]. A negentropy equal to zero indicates that y is Gaussian.

[4] $P_1(\mathbf{s}) = e^{-\lambda} \frac{\lambda^k}{k!}$, where λ was computed the average of products present in the market baskets, and k is the number of products present for a state $\mathbf{s}$.

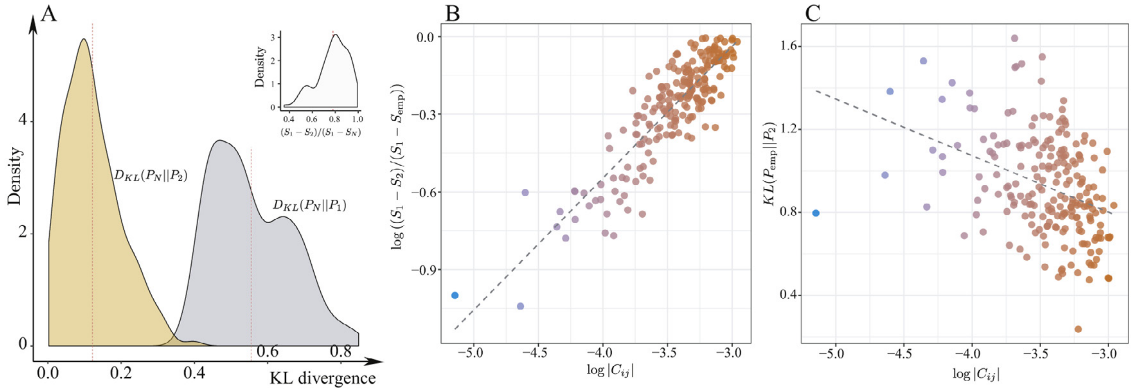

Fig. 5. **A**, Distribution of the Kullback–Leibler divergence for all sets of the M samples, between the empirical and independent model $D_{KL}(P_{emp} \parallel P_1)$, and between the empirical and pairwise model $D_{KL}(P_{emp} \parallel P_2)$; **B**, Comparing the magnitude of the two-body connections $\log(|C_{ij}|)$ with the log of the ratio I_2/I_N. The dashed line represents a fit between the two variables ($RMSE = 0.096$, $R^2 = 0.76$, $\beta = 0.509$). **C**, Comparing the magnitude of the two-body connections $\log(|C_{ij}|)$ with the Kullback–Leibler divergence between $P_2(\mathbf{s})$ and $P_{emp}(\mathbf{s})$ ($RMSE = 0.224$, $R^2 = 0.14$, $\beta = -0.272$).

when it comes to activations with a number of more than 5 products as shown in Fig. 3**C**. Despite the above, the pairwise distribution has a behavior more similar to that of the observed distribution than the independent model P_1. As shown in Fig. 5**A**, the Kullback–Leibler divergence $D_{KL}(P_{emp} \parallel P_2)$ and $D_{KL}(P_{emp} \parallel P_1)$ are well separated. In fact, the mean of $D_{KL}(P_{emp} \parallel P_2)$ is 0.122, while the mean of $D_{KL}(P_{emp} \parallel P_1)$ is 0.555.

Since the ratio I_2/I_N expresses the proportion of information that is attributed to pairwise interactions, it is expected that this measure has some relation to the existing correlations between pairs of C_{ij} products. This is precisely what Fig. 5**B** illustrates, in which the logarithm of these two quantities is compared. In this graph, the x-axis includes the averages of the existing correlations between all pairs of products for each sample. As the amount of correlation increases in absolute terms, so does the information explained by these second order interactions. Similarly, and verifying the previous result, Fig. 5**C** shows that as the level of correlation increases in the system, the divergence between the distribution of observed states and that of maximum entropy tends to decrease. The fact that there is variability in the divergence for different correlation values expresses evidence that higher-order effects are present in the collective purchasing system.

5. Application to a transactional sample dataset

Recently, it has been found that, similarly to the stock market, the aggregated purchasing behavior of retail consumers tends to form its own hierarchical structure [28]. The construction of minimum spanning tree (MST) using the Prim's algorithm, based on the correlations between product activity, is a good alternative for discovering the hierarchical structure. From a graph oriented perspective, every element of the system is a node and edges between them are obtained from a transformation of a distance measurements, as for example, a correlation matrix. From this graph, a MST is found which reduces the space from $N(N-1)/2$ correlations to $(N-1)$ edges [29]. Next, from the MST, it is possible to construct a hierarchical tree by means of ultrametric distances [30], see [1] for a detailed description of the process.

As the coupling parameters describes the non-linear interaction between spins of the system, it is possible to use the couplings matrix $\{J_{ik}\}$ to find the hierarchical organization of the system, and relate the to their energy levels. The couplings matrix is a symmetric $N \times N$ matrix being N the number of spins or products. However, the values of the couplings matrix must be transformed in such a way that positive coupling values represent closeness between two products, while negative values indicate larger distance. The MST can be found using Prim's algorithm using as an input the transformed values of the coupling matrix, $\{d_{ik}\}$. Then, ultrametric distances d^{ultr}_{ik} are found. The ultrametric distance between node i and k will be the maximum distance found in the path between the two nodes through the MST [30].

To exemplify the relationship between the hierarchical structure and the energy calculated on the basis of the parameters of the Ising distribution, we take a sample of any of the 250. We calculate the MST using the couplings inferred with Boltzmann learning[5] and for each of the possible states of this collective purchasing system, we calculate their energies and ultrametric distances d^{ultr}. Fig. 6 shows the corresponding MST and the dendrogram using the ultrametric distances between each product.

It is easy to see that for instance, product subcategories 80 (vitaminized pasta) and 7 (vegetable oil) are in the same hierarchy and represent by itself a group that is at the same ultrametric distance. Likewise, both have a strong positive coupling, suggesting that these two products tend to be present (or absent) simultaneously in market baskets. Then, at

[5] An appropriate transformation was used to represent positive coupling as closeness between two products, and negative values as farness.

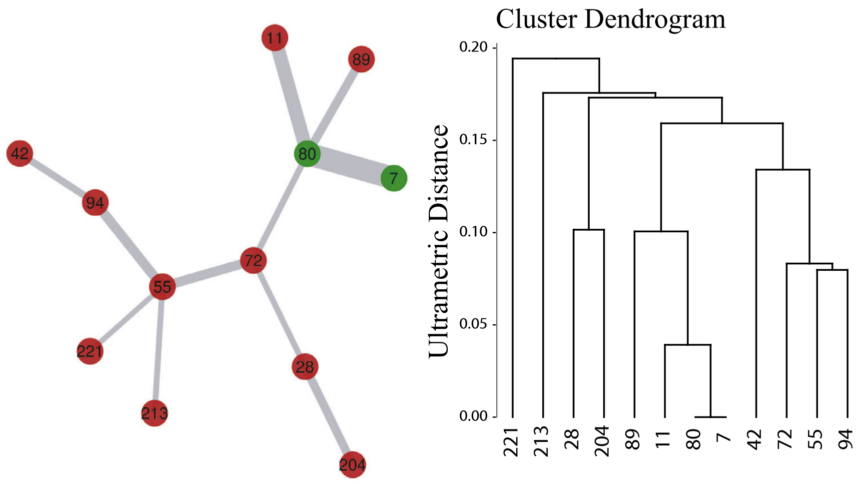

Fig. 6. Resulting MST using a proper transformation of coupling parameters inferred from the sample data. A thicker edge between two nodes indicates greater closeness. Green nodes indicates products with $h_i < 0$ and the red ones with $h_i > 0$. The ultrametric distances in the dendrogram are subtracted by the minimum of those distances to improve the visualization.

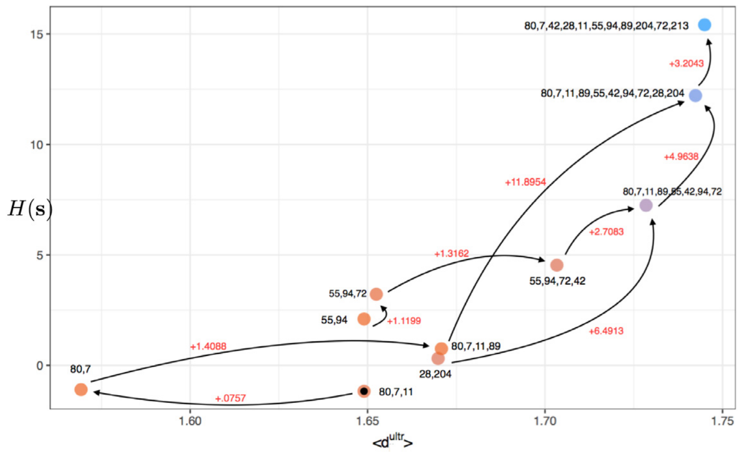

Fig. 7. Energy changes when passing from one hierarchical structure to another. The red numbers represent the energy needed to move from one state to another, indicated by arrows. The black numbers identify the active products in each state.

the next level of the hierarchy, subcategory 11 is added to the group, which is at the nearest distance. In fact, in the MST, elements 11 (rice grade 2) and 80 are also very close to each other, through a positive coupling.

We consider that, according to the hierarchical structure produced by the ultrametric distances, each group of products is equivalent to a particular state of the system. Thus, it is possible to know the energy requirements of each group (represented by the state of the active products) to overcome the ultrametric distance barriers. In other words, we want to know what is the increase in utility required to meet market baskets with products from different hierarchical groups.

In Fig. 7, the black dot represents the global minimum. In this state there are three active products: 80, 7, and 11 and all three are at an ultrametric distance of $d^{ultr} = 1.648$. These products together are in the same hierarchical group. An energy increase of $\delta E = 0.0757$ produces a jump to a new hierarchy and state given by products 80 and 7, that is, product 11 is left out of the market basket. Again, an energy increase of $\delta E = 1.4088$ overcomes a barrier of ultrametric distances to reach a new state consisting of products 80, 7, 11 and 89 (canned mackerel), which are at the same distance of $d^{ultr} = 1.669$. From this state, a substantial increase of energy of $\delta E = 11.8954$ allows to overcome two ultrametric distance barriers, reaching the hierarchical group composed by products 80, 7, 11, 89, 55, 42, 94, 72, 28, 204. The energy needed to reach this state is high, so it is a very unlikely state in the system. However, it is interesting to note that this state can be reached from another starting point with energy increments starting from the group/state 55 and 94 or 28 and 204.

6. Conclusions

This study has shown that it is possible to describe transactional data using the maximum entropy principle, in which the aggregate purchasing system of a multitude of transactions can be analyzed with a pairwise Ising model. Using Boltzmann learning, it was possible to find the field parameters of each product and the couplings between pairs of products that allows to parameterize the distribution of market baskets. The maximum entropy distribution $P(\mathbf{s})$ is consistent with the structure of correlations between product pairs and activations (presence) of the products in the market baskets.

The inferred Ising model for different samples, indicated that second-order interactions explain on average 77.9% of the information contained which implies that the aggregate buying behavior of thousands of consumers emerges from interactions between pairs of product items.

The Ising pairwise model is useful to understand certain properties of the system. For example, for certain products, field energy dominates the energy provided by interaction with other products, while on the other hand, there are products where interaction dominates the activation energy. Even when different samples have been used, the identification of jammed states or states of minimum energy revealed that there are products that consistently are present in low energy configurations. This is evidence of consistency between model estimates with different samples and different combinations of product groups.

Furthermore, the use of the couplings to obtain a network representation of the purchasing system and the hierarchical structure that emerges from the topology of that network, is extremely interesting because it allows to understand the link that exists between the energy of a state and its revealed hierarchy. However, there are also higher order interactions that explain the divergence of the Ising distribution from the empirical one, in basket sizes greater than five subcategories of products. In addition, we visualize that the inferred distribution has practical applications, useful to find states of minimum energy or *market baskets* that represent a set of products that deliver high utility and high frequency purchasing patterns.

Acknowledgments

The authors would like to thank CONICYT-Chile under grant Fondecyt 11160072 (M.A.V.) and Basal(CONICYT)-CMM, Fondecyt 1180706 (G.A.R.) for financially supporting this research.

References

[1] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1) (1999) 193–197.
[2] G. Bonanno, N. Vandewalle, R.N. Mantegna, Taxonomy of stock market indices, Phys. Rev. E 62 (6) (2000) R7615.
[3] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertiész, Dynamic asset trees and portfolio analysis, Eur. Phys. J. B 30 (3) (2002) 285–288.
[4] E. Scalas, The application of continuous-time random walks in finance and economics, Physica A 362 (2) (2006) 225–239.
[5] Y. Shapira, Y. Berman, E. Ben-Jacob, Modelling the short term herding behaviour of stock markets, New J. Phys. 16 (5) (2014) 053040.
[6] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, H.E. Stanley, Random matrix approach to cross correlations in financial data, Phys. Rev. E 65 (6) (2002) 066126.
[7] S. Maslov, M. Mills, Price fluctuations from the order book perspectiveempirical facts and a simple model, Physica A 299 (1–2) (2001) 234–246.
[8] I. Zovko, J.D. Farmer, et al., The power of patience: a behavioural regularity in limit-order placement, Quant. Finance 2 (5) (2002) 387–392.
[9] J.-P. Bouchaud, M. Mézard, M. Potters, et al., Statistical properties of stock order books: empirical results and models, Quant. Finance 2 (4) (2002) 251–256.
[10] T. Raeder, N.V. Chawla, Modeling a store, in: Advances in Social Network Analysis and Mining, IEEE, 2009, pp. 164—169.
[11] T. Raeder, N.V. Chawla, Market basket analysis with networks, Soc. Netw. Anal. Min. 1 (2) (2011) 97–113, http://dx.doi.org/10.1007/s13278-010-0003-7.
[12] A.-L. Barabási, Scale-free networks: a decade and beyond, Science 325 (5939) (2009) 412–413.
[13] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (4) (1957) 620.
[14] E.T. Jaynes, Information theory and statistical mechanics. ii, Phys. Rev. 108 (2) (1957) 171.
[15] E. Schneidman, M.J. Berry II, R. Segev, W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature 440 (7087) (2006) 1007.
[16] G. Tkačik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, M.J. Berry II, Searching for collective behavior in a large network of sensory neurons, PLoS Comput. Biol. 10 (1) (2014) e1003408.
[17] D. Fraiman, P. Balenzuela, J. Foss, D.R. Chialvo, Ising-like dynamics in large-scale functional brain networks, Phys. Rev. E 79 (6) (2009) 061922.
[18] E. Ganmor, R. Segev, E. Schneidman, The architecture of functional interaction networks in the retina, J. Neurosci. 31 (8) (2011) 3044–3054.
[19] H.C. Nguyen, R. Zecchina, J. Berg, Inverse statistical problems: from the inverse ising problem to data science, Adv. Phys. 66 (3) (2017) 197–261.
[20] T. Bury, Market structure explained by pairwise interactions, Physica A 392 (6) (2013) 1375–1385.
[21] H. Haken, J. Portugali, Information and selforganization: A unifying approach and applications, Entropy 18 (6) (2016) 197.
[22] G.E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800.
[23] D. Landau, K. Binder, A guide to Monte Carlo simulations in Statistical Physics, Cambridge University Press, 2014.
[24] L. Paninski, Estimation of entropy and mutual information, Neural Comput. 15 (6) (2003) 1191–1253.
[25] M. Mézard, G. Parisi, M. Virasoro, Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications, Vol. 9, World Scientific Publishing Company, 1987.
[26] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[27] E. Schneidman, S. Still, M.J. Berry, W. Bialek, et al., Network information and connected correlations, Phys. Rev. Lett. 91 (23) (2003) 238701.
[28] M.A. Valle, G.A. Ruz, R. Morrás, Market basket analysis: Complementing association rules with minimum spanning trees, Expert Syst. Appl. 97 (2018) 146–162.
[29] C.G. Gilmore, B.M. Lucey, M. Boscia, An ever-closer union? Examining the evolution of linkages of european equity markets via minimum spanning trees, Physica A 387 (25) (2008) 6319–6329.
[30] R. Rammal, G. Toulouse, M.A. Virasoro, Ultrametricity for physicists, Rev. Modern Phys. 58 (3) (1986) 765.