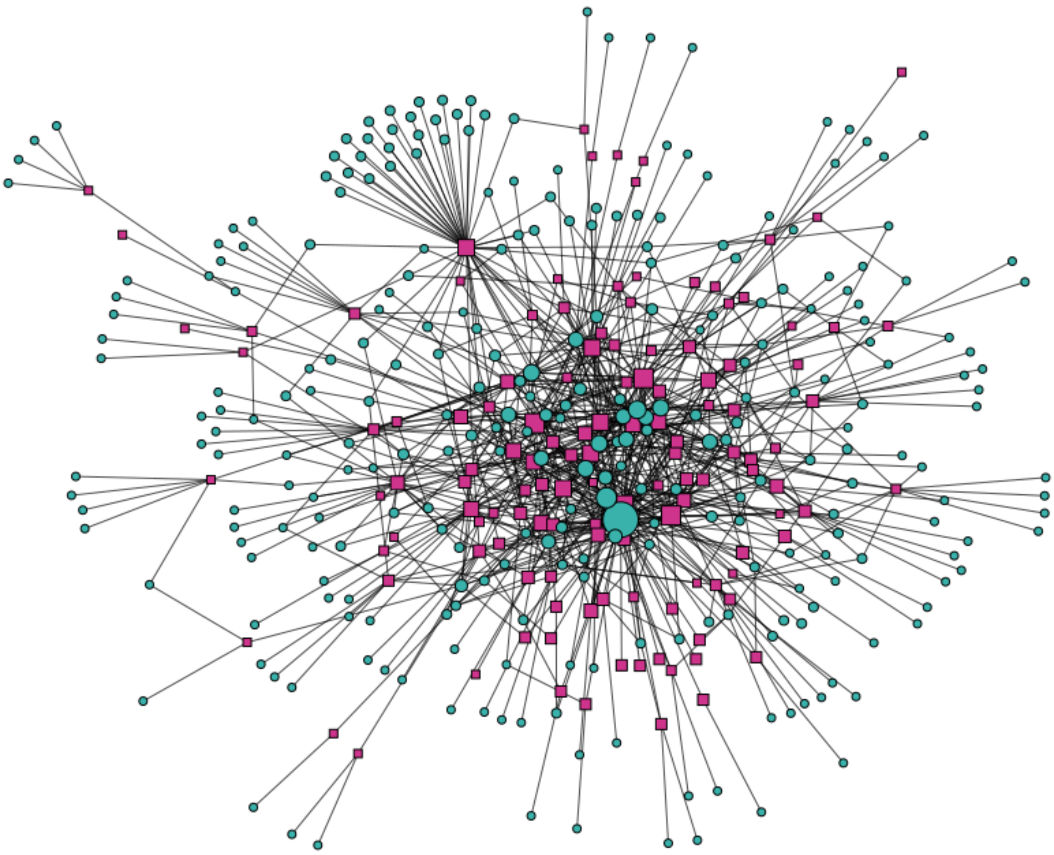
Engineering and Physical Sciences Research Council

InFoMM Industrially Focused Mathematical Modelling

# EPSRC Centre for Doctoral Training in Industrially Focused Mathematical Modelling



# Community structure in product-purchase networks


**A. Roxana Pamfil**


UNIVERSITY OF OXFORD

dunnhumby

# Contents

# 1 Introduction

## Shopping networks

With the introduction of store loyalty cards in recent decades, there has been an ever-growing body of data on shopping habits and preferences. Despite such rich data, understanding and predicting consumer behaviour remains a difficult problem. This is due to the myriad of factors that affect consumers' decisions (anything from mood to current weather), the inherent incompleteness of the data (as customers may frequent more than one retailer), and the ever-changing customer needs (such as increased preference for healthier foods).

Choices of weights in a shopping network include **item count** (how many bananas did this customer buy?) and a normalised variant called **item penetration** (what proportion of the items bought by this customer are bananas?).

Network science refers to the study of objects (**nodes**) and the connections between them (**edges**). Examples of networks include neurons interacting via synapses, people connected on an online social network, and cities linked by roads. A simple network model of shopping activity consists of customers connected to products that they purchased (see Figure 1). We call this structure a **product-purchase network** or a **shopping network**. Because some of these purchases occur in higher volumes than others, it is useful to add **weights** to the edges; edges that have larger weights correspond to stronger affinities between the respective customers and products.

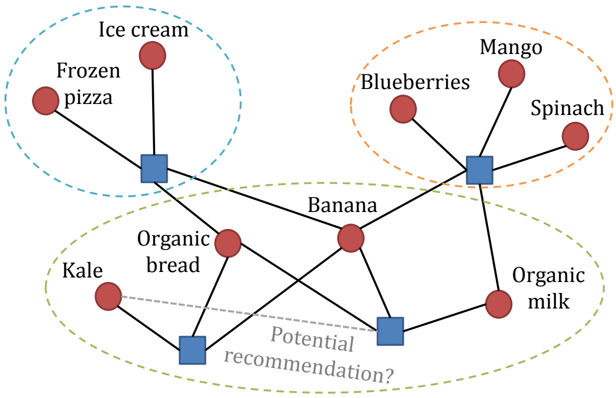



**Figure 1 – Example of a shopping network with three communities.**

## Community detection

There are many algorithms for community detection. Two popular approaches are maximising an objective function called **modularity** and performing statistical inference using **stochastic block models**.

Most real-world networks have a complex structure that is neither fully random nor fully regular, and characterising this structure is an important topic in network science that is traditionally known as **community detection** [2]. In qualitative terms, communities are sets of nodes that are densely connected with each other and sparsely connected with other nodes in a network. Examples of communities in empirical networks include left-leaning and right-leaning political blogs in the US, scientists who frequently co-author papers, and groups of friends in social networks. In a shopping network, communities reveal customers with similar preferences and the products they buy most. Detecting this structure is valuable for designing personalised recommendation systems and for other business insights.

## Data and network construction

dunnhumby provided access to "pseudonymised" transaction data from stores belonging to a major retailer in the UK. The data was pseudonymised for general research purposes (i.e., by replacing personally identifiable information with numeric IDs), such that no individual shoppers can be identified.

Constructing a network from transaction data requires defining a set of unique customers, a set of products, and a time period over which to aggregate purchases. We use 3-month time windows, and we typically weight the edges using **item penetration**.

See the PhD dissertation [3] for more details.

### Key contributions

By using mathematical models to perform **edge prediction** in a shopping network, we can address two business problems that are of interest to dunnhumby:

1. **[Recommendations]** What new items should we recommend to a given customer?
2. **[Targeting]** Which customers should we contact in a promotional campaign?

In Section 2, we discuss a validation of our targeting approach using historical data.

For many applications, it is useful to consider network models that incorporate additional features beyond a set of nodes linked together by edges. In Section 3, we summarise our findings for three such network representations, and we explain how these results are useful for analysing shopping networks.

### Glossary of terms

- **Network:** Mathematical representation of a system in which objects called **nodes** interact with each other via **edges**, typically in a pairwise fashion.
- **Community detection:** A process of assigning nodes in a network to cohesive groups called **communities**, such that nodes in the same community have more edges among them than to nodes in other communities.
- **Modularity maximisation:** A community-detection method based on ascribing a quantity called the **modularity** to any partition of a network into communities, then seeking a partition that has the largest possible modularity value.
- **Stochastic block model (SBM):** A statistical model of networks, typically used to **infer** community structure that is likely to have generated an observed network.
- **Edge prediction:** The process of using a model to predict new edges in a network (e.g., new connections between customers and products).

## 2 Network targeting in promotional campaigns

### Summary of approach

A **probability** is a number between 0 and 1 that reflects how likely it is for an event to happen. We use network models to estimate the probability that a customer will purchase a given product.

Figure 2 illustrates the key steps in identifying relevant customers to target in a promotional campaign (see also [4]). The first two steps, as explained in Section 1, consist of constructing a network from transaction data and performing community detection using a method such as modularity maximisation. The next step is to use a stochastic block model to estimate a set of **edge-propensity parameters** $\theta_{rs}$ that describe how nodes in different communities connect with each other. More precisely, the observed probability that a customer in community $r$ buys a product from community $s$ is proportional to $\theta_{rs}$. We now have a statistical network model of our data that makes it possible to calculate, for any customer and product, the probability of an edge existing between them. Sorting these purchase probabilities from largest to smallest produces a customer ranking that reflects their affinity to the target product. The top customers can then be sent coupons to elicit an initial purchase.

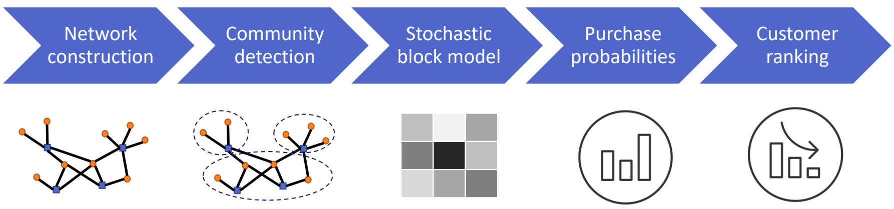


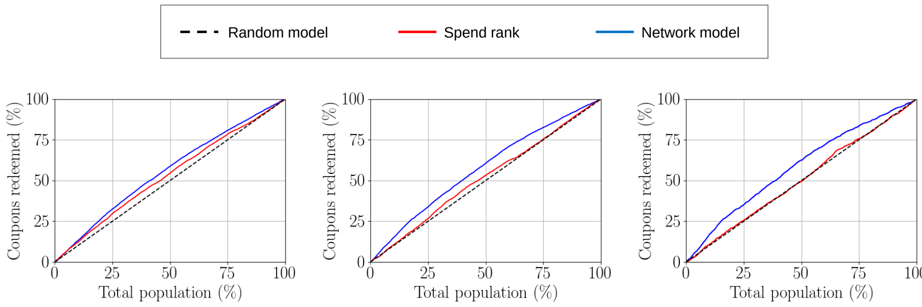

**Figure 2 – Key steps in a network-science approach to targeting.**

## Validation using historical data

To test how this method compares to simpler targeting approaches, we analysed data from a campaign in which a promotional offer for a specific brand of yoghurts was mailed to 100,000 customers. Using the approach from Figure 2, we rank these customers according to their affinities to the promotional yoghurts. If higher-ranked customers redeem more coupons than lower-ranked customers, then the corresponding ranking has predictive power, which suggests that the same approach can be useful to identify relevant customers for future campaigns. We use a ranking based on the customers' overall spend on yoghurts as a baseline for comparison.

In a **gains chart**, one plots the percentage of positive responses (in our case, the percentage of coupons redeemed) as a function of the population size (adding customers to the population one by one in the order given by the ranking).

We thus have two ways of ranking customers based on their affinity to the promotional yoghurts, and we compare these rankings in Figure 3 using **gains charts**. A ranking is predictive if there are more positive responses among the higher-ranked customers, and this corresponds to a curve that lies above the diagonal line in each chart.



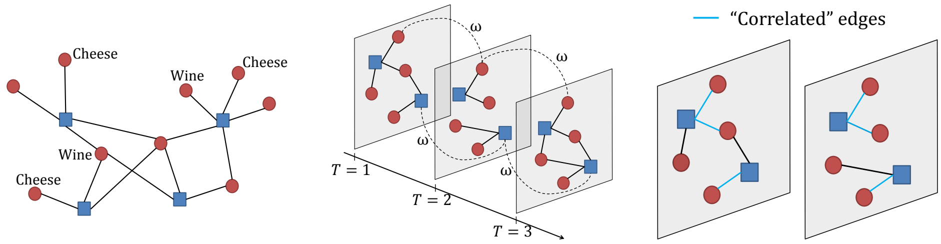

**(a) High-spend customers** **(b) Medium-spend customers** **(c) Low-spend customers**

**Figure 3 – Comparison of two ways of ranking customers that were contacted as part of a promotional campaign on yoghurts. Our industrial collaborator divided these customers into different segments based on how much they had spent on the promotional yoghurts prior to the campaign.**

The results show that our network model is better than category spend at identifying customers who are likely to redeem the coupon. This suggests that we can capture a customer's affinity for a product based on their similarity to customers within their community, and this works better than simply relying on a customer's historical purchases. It is particularly interesting that the performance gap is more pronounced among low-spend customers, who are less engaged with the brand on promotion.

# 3 Different network representations

In our work, we have considered several network representations beyond the most basic one. Three examples are **annotated networks**, **temporal networks** (which are a specific type of **multilayer network**), and **correlated networks** (see Figure 4).




**(a) Annotated network** **(b) Temporal network** **(c) Correlated network**

**Figure 4 – Illustrations of three network representations that we examine. (a) Annotated networks incorporate additional data, such as product categories, as labels on the nodes. (b) A temporal network, in which transactions from successive time periods belong to different layers, is one example of a multilayer network. (c) In a correlated network, edges that connect the same pair of nodes in multiple layers are no longer independent.**

## Annotated networks

Annotated networks include some additional **metadata** on the nodes, such as text labels or numbers. These annotations make it possible to incorporate additional domain-specific knowledge into a network model. In our simple models from Section 1, any two products are considered equally distinct: black grapes are no more similar to green grapes than they are to ice cream. Using product categories as annotations on the product nodes introduces evaluations of similarity between products.

Some algorithms provide **deterministic** community assignments (e.g., the "banana" node is in community 2), whereas others provide **probabilistic** assignments (e.g., the "banana" node is with 10% probability in community 1 and with 90% probability in community 2).

When using an algorithm to partition an annotated network into communities, those partitions that assign products from the same category to the same community become more likely. In other words, the algorithm favours partitions in which annotations better align with detected communities. We perform this task using a specialised stochastic block model (SBM). This approach returns **probabilistic community assignments**, meaning that each node can belong with some probability to each community.

We found that community detection in annotated networks leads to improved results, as it reduces the uncertainty around the correct community membership of product nodes (see Figure 5). An additional benefit is that the underlying SBM allows one to predict which customers would be most interested in a new product before it even appears on a shelf, just by considering the category of that product.

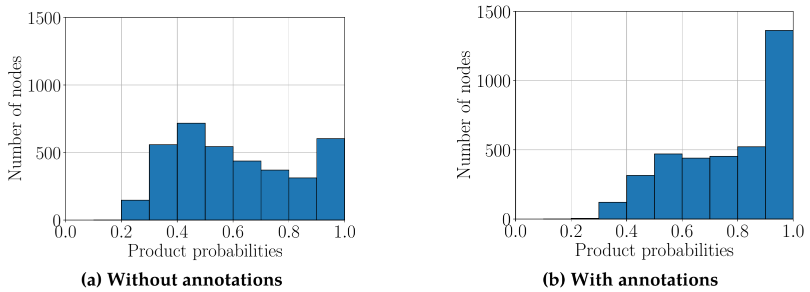



**Figure 5 – These histograms show, for all product nodes, the extent to which they belong to their best community (expressed as a probability). (a) Without annotations, there is significant uncertainty as the model assigns many products with just 30–60% probability to their best community. (b) With annotations, these probabilities get skewed towards 1.**

## Multilayer networks

It is possible to detect communities in multilayer networks by generalising single-layer modularity maximisation and SBMs. Our work in [1] uncovers an equivalence between these two methods under specific assumptions.

Many real-world data sets are better represented as a collection of interrelated networks rather than as a single graph. Because consumer preferences change over time, it is natural to use **temporal networks** (which are a specific type of **multilayer network**) to represent transactions over several time windows.

Detecting communities in multilayer networks is an active area of research. One popular method is to maximise a generalised version of the modularity objective function. Writing down multilayer modularity requires the specification of two types of "resolution" parameters, and choosing these appropriately is crucial for uncovering meaningful community structure in networks. One of our key results in this area is a method for inferring these parameter values in a principled way [1].

In Figure 6, we show a visualisation of community structure in a temporal shopping network with 12 layers, each corresponding to purchases during one 3-month time period. Each vertical slice of black rectangles depicts the communities in one layer, and the coloured areas between successive layers illustrate how nodes flow from one community to another across time. There are 3 large communities in each layer that are roughly constant in size. We found that the customer nodes in these communities correspond roughly to price-sensitive, mid-range, and high-end consumers. About 8% of the nodes change community membership between consecutive time periods, but we observe no broad shifts in shopping preferences in the data set that we analysed.

We have also studied another type of multilayer network called **multilevel networks**. For shopping networks, these provide a way to incorporate existing product hierarchies

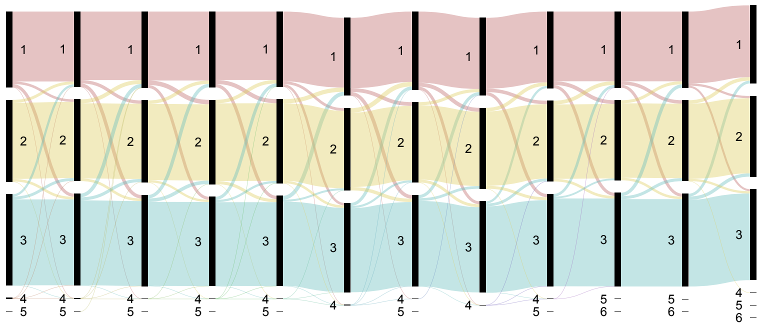



**Figure 6 – Visualisation of detected communities in a temporal shopping network.**

into the network structure. This is useful because the correct level of aggregation when constructing networks from transaction data is often not obvious from the start: should loose bananas be considered the same product as packed bananas, or should they be kept as separate nodes in the network? Our work provides a way to answer such questions and to perform community detection in other types of multilevel networks [3].

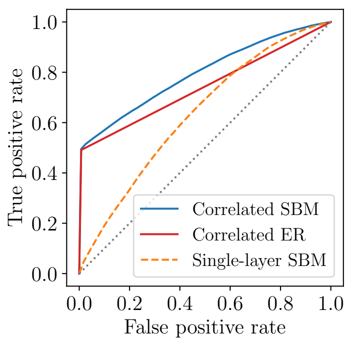
## Correlated networks

In temporal shopping networks, there is significant "edge persistence": customers tend to buy not just the same types of products over time (e.g., cereal), but specific brands within those wider categories (e.g., Cheerios). To capture these patterns, we introduce new models of **correlated networks** in which we no longer assume that edges in different network layers are independent. For instance, if a customer purchases a product in one time period, it becomes more likely that the same purchase occurs in a subsequent time period, compared to the case where no such previous purchase exists.

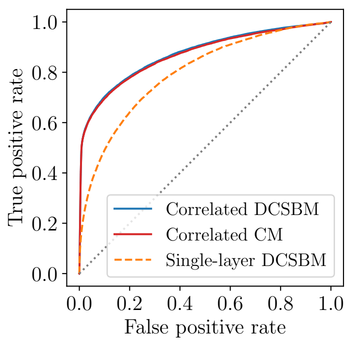
> An **ROC curve** is a standard tool for visualising the performance of a model that makes binary predictions (e.g., whether a customer purchases a product or not).

> The **degree** of a node in a network is its number of connections. Models with degree correction account for the fact that degrees tend to vary significantly in a network.

A key use of our correlated network models is to improve edge-prediction accuracy. To confirm this, consider purchases in two successive 3-month time periods. We aim to predict 20% of the observations in the second time period (i.e., whether a purchase occurs or not) by fitting a model using the remaining 80% of the data, alongside the observations from the first time period. In Figure 7, we use **ROC curves** to illustrate the predictive performance of four different correlated models, which we compare to two single-layer baselines that do not use information from the first time period. Curves that are further above the dotted diagonal lines in each panel indicate higher accuracy for the corresponding models. We see that correlated models significantly outperform their single-layer counterparts. This work has implications not just for consumer-behaviour applications, but also in fields like biology, where aggregating different (correlated) data sources is important for overcoming the high level of noise that is present in the data.




**(a) Models without degree correction**




**(b) Models with degree correction**

**Figure 7 – ROC curves for the edge-prediction task on a temporal shopping network. The dotted diagonal line in each plot indicates the expected ROC curve for random guessing. All other curves lie above this diagonal line, suggesting that they have predictive power. In both panels, correlated models outperform their single-layer counterparts.**

# 4 Discussion, conclusions & recommendations

Networks offer a natural representation of grocery purchases, which represent interactions between customers and products. Edge prediction is a standard task in network science that enables us to address two business problems that are important to dunnhumby: product recommendations and promotion targeting. Network models can be extended relatively easily to more complicated scenarios that include, for example, product-category metadata and temporal information such as a customer's repeated purchasing of some products.

We recommend that dunnhumby test the targeting approach from Section 2 in a live promotional campaign, to complement our retrospective analysis of a campaign for yoghurts. One advantage of networks is that they work well in "sparse" settings, where other methods might struggle for lack of sufficient purchasing data. This points to opportunities to use community-based recommendations beyond grocery, for slow-moving goods. Lastly, our work suggests ways of addressing the "cold-start problem" of recommendation systems. By performing edge prediction in either annotated or multilevel networks, it is possible to provide introductory offers to customers who are most likely to buy a newly-launched product, despite there being no historical purchasing data for that particular product.

The beauty of network science is that methods are rarely application-dependent. As a result, our findings in [3] are not specific to networks of customer and products. Instead, they can be useful for analysing community structure in any network across a variety of domains.

Rosie Prior, Academic Partnerships Manager at dunnhumby, said: "*Roxana's research delivered value to dunnhumby throughout the three years. Our data science teams appreciated her regular updates and companywide talks. As a Knowledge Transfer Ambassador, Roxana embedded her research findings at dunnhumby by giving a hands-on training course to all our data scientists and also writing a code library for them to use. She has supported the growth of network science at dunnhumby and has created an innovative solution for targeting relevant products to customers.*"

Jason O'Sullivan, Data Science Manager at dunnhumby, said: "*Roxana has made an incredibly valuable contribution to dunnhumby through her research. The research she has delivered has potential applications across multiple areas, with her primary focus being customer targeting, but also with potential applications to assortment and supply chain. More generally, her depth of knowledge in network science has enabled her to consult on multiple different pieces of work that have helped grow the depth of expertise within the business. Her work as a Knowledge Transfer Ambassador has created a real spark of network science understanding within the business that will lead to further and more extensive applications of network science within dunnhumby in the upcoming years.*"

# References


[1] A. R. Pamfil, S. D. Howison, R. Lambiotte, and M. A. Porter. Relating modularity maximization and stochastic block models in multilayer networks. *arXiv preprint arXiv:1804.01964*, 2018.

[2] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics Reports*, 659: 1–44, 2016.

[3] A. R. Pamfil. *Communities in Annotated, Multilayer, and Correlated Networks*. PhD thesis, University of Oxford, 2018.

[4] A. R. Pamfil. Network models for recommender systems. `https://medium.com/dunnhumby-data-science-engineering/network-models-for-recommender-systems-7f0d6d210ccf`, 2019. Accessed: 2019-05-14.