MorphoDyn(AGD) or how to visualize dynamics in high dimensions

MorphoDyn® or how to visualize dynamics in high dimensions

AGD - Global Pricing Team, May 2015

Digital Insurance market is getting more complex every day and pricing more sophisticated, agile and data-driven. To quickly extract key information from complex and high-dimensional structures, the Global Pricing Team developed the MorphoDyn approach.

1. Introduction

Customers' characteristics are composed of many features with various types of interactions between them. Thus, it is tricky to answer questions such as: Are the potential clients visiting our website today different from the ones that came one year ago? How are the new clients' profiles evolving month by month? Did the vehicles' characteristics of our clients change?

Meanwhile, the insurance market is also evolving and to take advantage of any changes, it is crucial to identify with whom AXA is competing and capture the impact of a tariff change on the new business profiles.

MorphoDyn is designed to help answer these questions as it displays practical and easy-to-understand visualization of high-dimensional structures.

2. Algorithms behind the approach

The core idea of MorphoDyn is to visualize the structure of the correlations, both in a static and dynamic way.

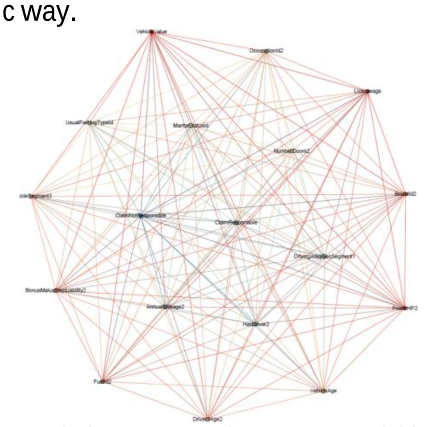


Figure 1: Correlations structure between 18 variables where it is hard to identify the variables of interest.

Defining indicators to follow the correlations between each pair of variables leads to too much information which can obscure the important signals.

The Minimum Spanning Tree (MST) is a graph that covers all nodes without a loop and enables dimension reduction. With the MST, the *n* variables are only linked by *n-1* edges and the essential structure of the dependence is kept.

To visualize the structure in an obvious manner:

- each node stands for a single client characteristic or competitor depending on the scenario
- the node's size identifies its importance compared to the other nodes and is based on the *PageRank Score* algorithm (standard measurement nodes importance in networks [4])
- the node's color describes how the final tree is compartmentalized into sub-networks, by reflecting classes of modularity (automatic clustering algorithm [5])
- the layout is intuitive as it is based on Physics theories of attraction and repulsion (Force Atlas Layout algorithm) and the distance between the nodes is likely to be small when they are correlated.

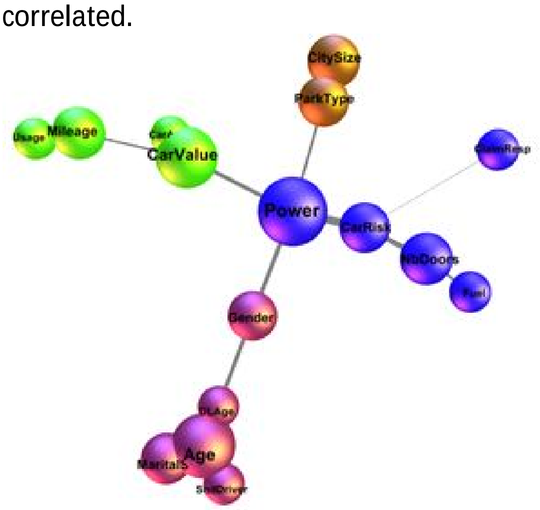


Figure 2: Visualization of Client Profile with MorphoDyn approach. Similar variables appear in the same sub-network i.e. Drivers variables (pink), Geographical info (gold). The key variable here is a vehicle variable: Power, whilst Gender links the vehicle's information to the driver's information.

3. Optimal time scale for dynamics follow-up

To make sure that MorphoDyn will capture relevant changes in the correlation structure, the proper time scale must be identified. Indeed, if the measurement time is too short, variables of interest will appear to evolve but this may just be due to non-reproducible measurement noise. If the measurement time is too long, important changes can go unnoticed, embedding the change variation in the information.

A meaningful way to follow-up the competitors' strategy is to identify the link and the dynamics between the competitors by studying the correlation between their prices. The underlying assumption is that, in the network of competitors, two players are close to each other if they have similar pricing strategies i.e. they tend to offer the same prices to the same customers.

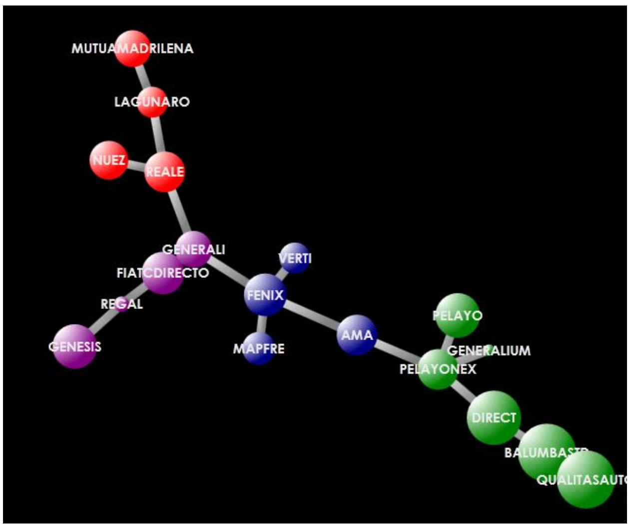
Technical snapshot: Random Matrix Theory

It is fundamental to understand the dynamics of the correlation matrix. Here, the approach consists of studying its eigenvector dynamics which bring a synthetic vision of the correlation matrix and relies on the Random Matrix Theory [3]. When trying to follow the evolution of an eigenvector, one faces a problem when eigenvectors hybridize due to a small change. An eigenvector hybridizes with the eigenvectors associated to close eigenvalues in this case. The solution is to study the stability of a subspace of several consecutive eigenvalues and more precisely, to study the stability of the first eigenspaces of two correlation matrices on two non-overlapping time windows. To do so, a fidelity distance is constructed on the rectangular matrix of overlaps of these two subspaces which is robust to small changes and noise.

The conclusion of the analysis is that the optimal observation period is two weeks for the Spanish market and the Italian market whilst, for the French market, the time scale corresponds to a one month period.

4. Applications

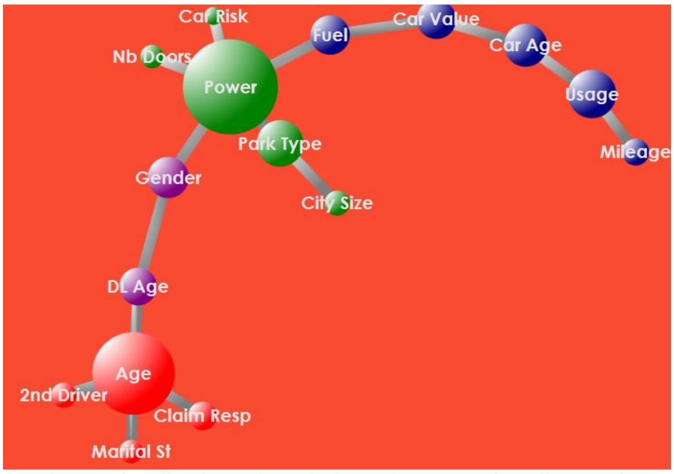
Dynamic view of an AGD entity market evolution



Movie 1: Visualization of competitors' network of an AGD entity by 2-week period throughout February 2014 to March 2015.

MorphoDyn helps to visualize complex market relations at a glance and identify the position of the competitors relatively to each other.

Dynamic evolution of clients profiles



Movie 2: Visualization of evolution of AGD entity clients who choose the Partial Comprehensive package (observations by quarter, moving by a monthly path)

Over an 18 month period, MorphoDyn detected and confirmed substantial impacts of major changes from the full review of GLM tariff (Mar. 2014), the commercial premium and/or underwriting rules changes (Jul. 2014 - Sep. 2014 & Nov. 2014 & Mar. 2015) and the launch of another AXA product (Jan. 2015).

5. Cases studies in practice in AGD entities

MorphoDyn is a powerful approach which already helps to:

- confirm that a specific price change impacts the clients mix (Spain)
- spot the differences between the prospects and the new clients profiles (Korea) – as a detection of antiselection or a deliberate strategy effect
- describe the differences between the clients of two companies (France)
- identify the competitors' network and the impact of our changes in our positioning (Spain)
- Grasp underlying structures between variables during initial analysis of a new database.

Major References:

[1] EMMA Initiative, Correlation and dependence measures: Academic literature and selected financial applications.

[2] D. Lautier; F. Raynaud (2011). Systemic risk in energy derivative markets: a graph-theory analysis.

[3] R. Allez and J.P. Bouchaud. Eigenvector dynamics: general theory and some applications. Arxiv preprint arXiv:1203.6228, 2012.

[4] L. Page, S. Brin, R. Motwani, T. Winograd (1999). The PageRank citation ranking: Bringing order to the Web.

[5] V. D.Blondel; J.L. Guillaume; R. Lambiotte; E. Lefebvre (2008). Fast unfolding of communities in large networks