Pairwise influences in dynamic choice: network-based model and application

Stefano Nasini\textsuperscript{a} and Victor Martínez-de-Albéniz\textsuperscript{b}

\textsuperscript{a}IESEG School of Management, LEM CNRS 9221; \textsuperscript{b}IESE Business School

ARTICLE HISTORY
Compiled September 27, 2019

ABSTRACT
In this paper we study the problem of network discovery and influence propagation and propose an integrated approach for the analysis of lead-lag synchronization in multiple choices. Network models for the processes by which decisions propagate through social interaction have been studied before, but only a few consider unknown structures of interacting agents. In fact, while individual choices are typically observed, inferring individual influences – who influences who – from sequences of dynamic choices requires strong modeling assumptions on the cross-section dependencies of the observed panels. We propose a class of parametric models which extends the vector autoregression to the case of pairwise influences between individual choices over multiple items and supports the analysis of influence propagation. After uncovering a collection of theoretical properties (conditional moments, parameter sensitivity, identifiability and estimation), we provide an economic application to music broadcasting, where a set of songs are diffused over radio stations; we infer station-to-station influences based on the proposed methodology and assess the propagation effect of initial launching stations to maximize songs diffusion. Both on the theoretical and empirical side, the proposed approach connects fields which are traditionally treated as separated areas: the problem of network discovery and the one of influence propagation.

KEYWORDS
Network influences; Cross-sectional dependencies; Multidimensional panel data; Exponential family of distributions; Music broadcasting industry

V. Martínez-de-Albéniz’s research was supported in part by the European Research Council - ref. ERC-2011-StG 283300-REACTOPS and by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economía y Competitividad) - ref. ECO2014-59998-P.
1. Introduction

Decisions of different individuals over time exhibit pairwise associations in a wide range of economic contexts [8, 16, 17, 23, 27, 52]. Clear examples are demand-side economies of scale – where the attractiveness of a product increases as a function of the total number of consumers [13] – and also several classes of monopolistic competition – where product differentiation depends on the imitation between companies [28]. This condition covers a wide range of market structures where decision makers are influenced\(^1\) by the decisions of others through imitation, such as in the information technology, apparel retailing or music broadcasting industries.

The availability of large social and economic network data sets over the past three decades has made possible to study social and economic influences like never before. Having said that, two complementary viewpoints have characterized the analysis of network influences.

- Building on the statistical hypothesis of Granger causality [34, 35], influence is invoked when sequences of individual decisions are systematically anticipated by the ones of others [6, 22, 67]. The aim is to disclose synchronized lagged similarities in dynamic choices, from which pairwise influences are empirically inferred.

- Relying on a known structure of interacting agents (often observed from a network data set), network analysts have focused on the inverse problem of network diffusion from a starting condition [5, 36, 45, 48, 57, 66], where the multiplier effect that influence might produce results in amplifying or shrinking the diffusion of choices among the agents.

Traditionally, the scientific communities working in the problem of network discovery and influence propagation rarely intersect. This justifies the scarcity of theoretical and empirical literature focusing on the complementarity and duality between both problems.

Consistently with these viewpoints, this paper proposes a novel parametric approach — we call it Pairwise Influences in Dynamic Choice models (PIDC models, from now on) — which provides a unified modeling framework for both purposes.

On the one hand, as described in Section 3, the PIDC class of models builds on the observation of a three-dimensional panel \( \{ x_{ist} \mid i \in I, s \in S, t \in T \} \), namely the realization of a triple-index process defined on a suitable probability space (where \( i \) is the item dimension, \( s \) is the individual dimension, \( t \) is the time dimension) and infers network patterns from cross-section dependencies in sequential decisions of mult-

\(^1\)The term influence is here adopted in a comprehensive sense to generalize the ones of imitation, and spillover, which are used in more particular context, such as social psychology and microeconomics respectively.
multiple individuals (companies, customers, decision makers) over multiple items (products, ideas, business outcomes). With the aim of assessing the hypothesis of Granger causality between the dynamic choices of pairs of individuals on items, the proposed methodology extends the well-known vector autoregressions [49, 59] and the recently proposed network vector autoregressions [67] to the case of discrete choices, based on the general distribution form of exponential random models [56].

On the other hand, as shown in Section 5, the PIDC class of models has straightforward implications for the influence maximization problem. Thoroughly, it supports the optimal selection of initial individuals to diffuse new choices, in line with the classical approaches of [36] and [57]. On this respect, the advantage of the PIDC class of models over existing network influence models for this purpose is the possibility to support a large variety of model specifications, which can be either driven by empirical observations (based on the network discovery side of the PIDC framework) or consistent with behavioral theories [7, 17, 23]. These model specifications integrate and extend the modularity patterns of the linear thresholds model [36, 57] and the independent cascade model [32], by allowing for submodular, linear and supermodular influence propagations.

Building on this twofold modeling framework, we uncover a variety of statistical properties (conditional moments, parameter sensitivity, identifiability and estimation) and provide a motivating business case in which the PIDC class of model can capture the diffusion of songs across broadcasting stations in the UK.2 Our data contains the number of plays of each song (corresponding to the item dimension of the panel) over many broadcasting stations (corresponding to the individual dimension of the panel). These bring to the market highly correlated program schedules – i.e., choices of songs with high cross-section dependencies arising as a result of pairwise imitations. Concretely, Figure 1 illustrates the weekly number of plays in the UK of two popular songs from Bruno Mars and Figure 2 restricts this graphical representation to the number of plays of the same songs in two specific stations: BBC Radio 1 Xtra and BBC Radio 2. Song plays exhibit a dynamics which resembles the short life cycle of fashion goods: the demand of songs evolves on a time window in which their popularity increases shortly after launch and then decreases.3 Contextually, stations follow an synchronization pattern where BBC Radio 2 might be seen as anticipating BBC Radio 1 Xtra (or the former imitating the latter).

2 Analogous business settings have been recently analyzed by [21] and by [65], from the viewpoint of the applied (bottom-up) statistical modeling.

3 When presenting our modeling framework, this time effect of song diffusion has been treated as an exogenous social trend (possibly due to an appetite for novelty, or to a planned marketing strategy), which affects broadcasting decisions without being affected by them [38]. This simplifying assumption is made explicit in Section 3.2 and allows us to estimate the shape (life cycle) of this exogenous social trend and its influence on each station.
In an effort to uncover the influence of one station on another and their ability of conditioning the future choices of other players in the market, the twofold orientation of the PIDC class of model allows (i) estimating the strength of station-to-station dependencies and (ii) optimizing the introduction decisions in the music broadcasting industry. In fact, with knowledge of the influence structure, music producers can better select which station to choose for first diffusing their products, as influence effects between broadcasting companies are determinant to maximize the diffusion of a song in the first weeks after launch. Conversely, our results show that these effects fade away with time as the song reaches the entire network, supporting the fact that short-run network effects are mitigated by long-run fashion trends.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 provides a detailed description of the proposed exponential random model for network discovery and influence propagation, along with the identifiability of parameters estimation. The empirical application to the context of music broadcasting is presented in Section 4. In Section 5 we examine the influence maximization under different specifications of the PIDC class of models and describe how the estimated model can help decision-makers better choose the stations where songs are released first. Section 6 concludes. The mathematical proofs of propositions are reported in Appendix A. An enlarged version of this manuscript is online available as a working
2. Literature review

This paper is connected to several streams of literature, which are for convenience grouped in three main fields: Statistical models of dynamic choices, Stochastic models for network diffusion and propagation and Network auto-regression models.

2.0.0.1. Statistical models of dynamic choices. Choice models consist in the probabilistic selection from sets of mutually exclusive alternatives [33, 44, 60]. In such context, multidimensional panel data generally appear as sequences of multiple choices by fixed collections of individuals [3, 31, 50].

Multinomial logit models are typically adopted as a referential framework for individual choices [43], which can be easily generalized to cases of infinite alternative by the Poisson asymptotic behavior [58], or properly parameterized to account for time-dependent properties in dynamic choice settings and correlated individual decisions [37, 61]. In fact, a common issue when studying dynamic choices of multiple individuals over a set of items concerns the presence of cross-sectional dependencies either in the item or in the individual dimension of the panel [4, 22, 47].

In some cases, when interactions between spatially distributed consumers, retailers or manufacturers are specified, spatial autoregressive processes have also been considered as models of panel data [48]. Besides spatial proximity, cross-correlations of errors could arise as a result of interactions within socioeconomic networks. For example, [5] estimate influences and spill-overs in an educational context by internalizing the within-group similarities of individual choices. Analogously, [25] propose an econometric approach for the identification of social interaction in multivariate choices and provide an empirical application to students’ academic performance.

Despite the sparse contributions in the econometric literature, systematic studies of such a structural analysis can only be encountered in the random graph and complex network literature, as discussed in the next paragraphs.

2.0.0.2. Stochastic models for network diffusion and propagation. Random models of network formation infer connection structures from purely stochastic processes where expected connections might depend on exogenous individual choices or outcomes [19, 51]. By contrast, traditional modeling approaches for network influence, diffusion and propagation assume fixed network structures and focus on the problem of studying the diffusion pattern from a given starting condition. They reframe the classical diffusion model by [9] (and subsequent extensions [10, 62]), analyzing the
spread of diffusion over a fixed population, without explicitly taking into consideration connections at individual level.

More than forty years ago, the DeGroot learning model [26] was one of the first approaches in this area. His basic intuition was to design the choice dynamics of a node from period to period by averaging out the choices of its neighbors.

Contextually, [36] and [57] investigated a diffusion model – called the linear thresholds model – where binary states of nodes (i.e., active versus inactive) is dynamically updated with a probability representing the total proportion of active neighbors. A similar approach has been proposed in the context of marketing by [32], where any active nodes are capable of activating its neighbors with a fixed independent probability. This is called the independent cascade model. The linear thresholds model and the independent cascade model are among the most widely studied mathematical frameworks for diffusion of influence taking place over the edges from node to node. [41] showed that influence maximization problems (the optimal selection of initial nodes in the network) under the linear thresholds model and the independent cascade model can be viewed as general cases of set cover problems, thus NP-complete.

Despite their large range of applicability, these models require a known structure of pairwise connection, which is not available in most empirical analysis. There are alternative approaches to deal with diffusion and propagation, based on pure simulation schemes, such as multi-agent systems [42, 55], though they do not allow to statistically obtain inferential insights about the influence structure from empirical observations.

2.0.0.3. Network auto-regression models. With a view to statistical inference and empirical analysis, the PIDC class of models can be included in the list of regression-like modeling approaches which internalize the cross-section dependencies between statistical unites based on a specified structure of network proximity.

In this respect, an early settlement of the problem has been addressed by [11] who considered a joint Markov random field for binary data, paving the way to the spatial auto-logistic regression. Some years later [29] presented the a generalization of the spatial effects linear model and the linear model with spatial error terms. In the same year [30] provided a simulation study of the consequences of ignoring the network autocorrelated disturbance.

Perhaps owing to the appeal of the models simple and direct specification of dependences, network auto-regression models have been widely applied in the context of individual choices. For instance, [63] study the link between student relations and their performances at university, building on the hypothesis of a social influence mechanism

---

4Moreover, none of these models allows for the simultaneous inclusion of multidimensional choices over a set of items whose attractiveness is non-homogeneous in time, as for music songs.
where individuals adjust their own behaviors to those of others with whom they are connected. Similarly, [12] investigate the impact of diversification and disintermediation strategies in the evolution of financial trading networks using a specialized version of the network auto-logistic regression.

As an alternative made necessary by our specific data setting, the versatility of PIDC class of models (in comparison to classical auto-regression and auto-logistic approaches) embraces a broader collection of exponential random models which allows capturing different forms of influences induced from well-known statistical families.

3. Model definition and specification

A statistical model for influence discovery is presented in this section, building on the idea of dynamic choices in the music broadcasting industry. We consider radio stations (individual dimension of the panel) bringing to the market songs play-lists (item dimension of the panel) along a collection of periods (time dimension of the panel).

The following notation is considered throughout this paper, where vectors are denoted in boldface letters, whereas sets are denoted in calligraphic letters.

Sets:
- \( \mathcal{S} \) a set of individuals (radio stations);
- \( \mathcal{I} \) a set of items (songs);
- \( \mathcal{T} \) a set of time periods (weeks).

Choices variables:
- \( x_{sit} \) choice of the \( s^{th} \) individual over the \( i^{th} \) item (e.g., the number of plays of song \( i \) in station \( s \)) at time \( t \).
- \( \mathbf{x}_{s,t} = [x_{s1t} \ldots x_{s|\mathcal{I}|t}]^\top \) \(|\mathcal{I}|\)-dimensional choices of individual \( s \) at time \( t \);
- \( \mathbf{x}_{i,t} = [x_{1it} \ldots x_{|\mathcal{S}|it}]^\top \) \(|\mathcal{S}|\)-dimensional individual choices on item \( i \) at time \( t \);
- \( \mathbf{x}_{:,t} = [(\mathbf{x}_{1:,t})^\top \ldots (\mathbf{x}_{|\mathcal{S}|:,t})^\top]^\top \) \(|\mathcal{I}| \times |\mathcal{S}|\)-dimensional individual choices on items at time \( t \).

5As described in detail in Subsection 4.1, the empirical observation of stations play-lists supports the presence of pairwise lead-lag synchronization in the dynamic selection of songs. Influences are invoked between the broadcasting decisions of stations (individuals) over songs (items), motivating the need of a novel parametric model to characterize their structure in a probabilistic way.
Let \( h : \mathbb{R} \rightarrow \mathbb{R} \) be non-negative and non-decreasing and \( G : \mathbb{R}^2 \rightarrow \mathbb{R} \) be defined as \( G(x,y) = xg(y) \), where \( g \) is a non-decreasing real valued function.\(^6\) Sometimes the shorter notation \( h_{sit} \) and \( G_{ss'itt} \) is adopted instead of \( h(x_{sit}) \) and \( G(x_{sit}, x_{s'it(t-\ell)}) \), along with \( g_{s-t} = [g(x_{sit}) \ldots g(x_{s|\Gamma}|_{it})]^\top \) and \( g_{it} = [g(x_{1it}) \ldots g(x_{|S|it})]^\top \). Note that \( G_{ss'itt} = x_{sit} g(x_{s'it(t-\ell)}) \) quantifies the similarity between the choice of individual \( s' \) on item \( i \) at time \( t - \ell \) and one of individual \( s \) on the same item \( i \) at time \( t \).

We build a conditional model for the decision \( x_{sit} \) that the \( s^{th} \) individual takes with respect to the \( i^{th} \) item at time \( t \):

\[
\mathbb{P}(x_{sit} \mid x_{s,i:t-\tau_{\min}} \ldots x_{s,i:t-\tau_{\max}}) = \frac{1}{Z_{sit}(\Gamma)} \frac{1}{(h_{sit})^\psi} \exp \left( \psi \sum_{\ell=\tau_{\min}}^{\tau_{\max}} \sum_{s' \in S} \gamma_{\ell ss'} G_{ss'itt} \right), \tag{1}
\]

where \( \psi \) is a fixed constant and \( \gamma_{\ell ss'} \) is the effect of the choices of individual \( s' \) at time \( t - \ell \), on the ones of individual \( s \) at time \( t \), with \( (s,s') \in S \times S \). In the described application to the music industry, it captures a (possibly time-delayed) spill-over or imitation between couples of stations. For each \( \ell = \tau_{\min} \ldots \tau_{\max} \), we define \( \Gamma_{\ell} \in \mathbb{R}^{n \times n} \), whose \( (s,s') \) component corresponds to \( \gamma_{\ell ss'} \). The function \( Z_{sit}(\Gamma) \) is a normalizing constant for the conditional model (1) (also known as the partition function).

- \( h_{sit} \) and \( G_{ss'itt} \) are exogenous and given by the user for each application. In particular, \( g_{s-t} = [g(x_{sit}) \ldots g(x_{s|\Gamma}|_{it})]^\top \) and \( g_{it} = [g(x_{1it}) \ldots g(x_{|S|it})]^\top \). The specification of \( h \) and \( g \) is exogenous and given by the user for each application. In particular, \( g \) captures the functional form of the lead-lag dependency as a function of past choices.

\( ^6 \) The specification of \( h \) and \( g \) is exogenous and given by the user for each application. In particular, \( g \) captures the functional form of the lead-lag dependency as a function of past choices.

**Definition 3.1** (Three-dimensional panel joint probability). Let the individual decisions at the first \( \tau_{\max} \) periods \( x_{s,.1} \ldots x_{s,.\tau_{\max}} \) be known. The joint probability of observing a sequence of \( |S| \) multidimensional choices from period \( \tau_{\max} + 1 \) to period \( T \) is defined based on (1) by assuming conditional independence between individuals and items and applying the product rule:

\[
\mathbb{P}(x_{s,.\tau+1} \ldots x_{s,.T} \mid \Gamma) = \prod_{s \in S} \prod_{\ell \in I} \prod_{t \in T} \mathbb{P}(x_{sit} \mid x_{s,i:t-\tau_{\min}} \ldots x_{s,i:t-\tau_{\max}}) = \frac{1}{Z(\Gamma)} h_{\Pi}(x) \exp \left( \psi x^\top \Gamma g \right), \tag{2}
\]

where \( x = [(x_{s,.\tau+1})^\top (x_{s,.\tau+2})^\top \ldots (x_{s,.T})^\top]^\top \), the function \( h_{\Pi}(x) = \prod_{t \in T} \prod_{\ell \in I} \prod_{s \in S} (h_{sit})^{-\psi} \), the lag \( \tau_{\min} > 0 \), the function \( Z(\Gamma) \) is a normalizing constant.
The components of vector $g \in \mathbb{R}^{|I||S||T|}$ correspond to $g(x_{sit})$, sorted in lexicographic order with respect to individuals, items, and time periods.$^7$

The characterization of the sample space $X \subseteq \mathbb{R}^{|I||S||T|}$ of this multidimensional random variable can either include constraints on the play-list capacity, or allow for unrestricted choices. The latter case is assumed in the oncoming results, to avoid the inclusion of unnecessarily tedious notation.

**Assumption 1** (The underlying measure). *The underlying measure $h : \mathbb{R} \rightarrow \mathbb{R}$ in (1) is such that $h'(x)/h(x)$ is monotonically non-decreasing and positive and $h(0) \geq 1$, where $h'$ is the first order (discrete or continuous) derivative.*

Assumption 1 is verified by most of the well-known exponential random models, such as the Poisson and the Gaussian model.

We denote with $\mathcal{F}$ the so-defined family of distributions over three-dimensional panels with probability measure (2) and note that not all influence structures $\{\Gamma_\ell\}_{\ell=\tau_{\min}}^{\tau_{\max}}$ lead to a well-defined probability distribution, due to the sufficient statistic $G_{ss'\ell\ell't}$ not being integrable. The domain $\mathcal{D}(\mathcal{F})$ is the set of all $\tilde{\Gamma}$s which lead to a well-defined probability distribution, that is to say, $\mathcal{D}(\mathcal{F}) = \{\Gamma_{\tau_{\min}} \ldots \Gamma_{\tau_{\max}} | Z(\tilde{\Gamma}) < \infty\}$.

The sensitivity of PIDC models to the influence structure $\{\Gamma_\ell\}_{\ell=\tau_{\min}}^{\tau_{\max}}$ can be assessed based on general properties of the exponential family of distributions.

**Proposition 3.2.** *At each period $t > \tau_{\max}$, for each item $i \in I$ and individual $s \in S$, the conditional model (1) verifies*

$$
\frac{\partial}{\partial \gamma_{ss'}} \mathbb{E} \left[ x_{sit} \mid x_{s',i,t-\tau_{\min}} \ldots x_{s',i,t-\tau_{\max}} \right] = g(x_{s'i,t-t}) \mathbb{V} \left[ x_{sit} \mid x_{s',i,t-\tau_{\min}} \ldots x_{s',i,t-\tau_{\max}} \right],
$$

$$
(3)
$$

$^7$Note that while the structure of $\tilde{\Gamma}$ is not critical for the theoretical properties presented in this section, in the music industry application the dependency pattern encoded in $\tilde{\Gamma}$ assumes that previous choices of stations have an impact on future choices. However, that dependency does not happen at the item level, where the number of plays of a song are not related on the number of plays of others. As discussed in Section 4, this assumption has its empirical foundation in the size of the item dimension, implying negligible complementarity effects between songs.
where \( \mathbb{E}[\, . \,] \) and \( \mathbb{V}[\, . \,] \) are respectively the expectation and the variance operators.

As a supporting statement, this mirrors the dynamic expectations in the presence of cross-section dependencies, which result to be amplified or shrunk by the similarities between stations play-lists.

When the sample space \( \mathcal{X} \) is continuous, propositions 3.3, 3.4 and 3.5 allow studying the location of the conditional model (1) in closed-form expressions, as a function of the influence structure \( \tilde{\Gamma} \).

**Proposition 3.3.** The conditional distribution (1) is unimodal when \( h \) verifies Assumption 1.

**Proposition 3.4.** Consider the conditional distribution (1) and let \( H : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|} \) be defined as \( H(x_1, \ldots, x_n) = [h'(x_1)/h(x_1) \ldots h'(x_n)/h(x_n)]^\top \). As long as \( h \) verifies Assumption 1, we have

\[
\text{mod}[x_{\cdot,i,t} | x_{\cdot,i,t-\tau_{\min}} \ldots x_{\cdot,i,t-\tau_{\max}}] = H^{-1}\left( \sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_\ell g_{\cdot,i,t-\ell} - Z(\tilde{\Gamma})e \right)
\]

where \( \text{mod}[\, . \,] \) is the mode of the distribution.

**Proposition 3.5.** Let \( m_{\cdot\cdot\cdot,i} \) be the mode of the conditional model (1), as defined in Proposition 3.4, and define \( \kappa_{\cdot\cdot\cdot,i}(\psi) = \psi h(m_{\cdot\cdot\cdot,i})^\psi \left( h''(m_{\cdot\cdot\cdot,i}) - (h'(m_{\cdot\cdot\cdot,i}))^2 / h(m_{\cdot\cdot\cdot,i}) \right) \).
Based on the Laplace approximation, when \( \psi \) grows large we have

\[
\mathbb{E}[x_{\cdot\cdot\cdot,i} | x_{\cdot\cdot\cdot,i,t-\tau_{\min}} \ldots x_{\cdot\cdot\cdot,i,t-\tau_{\max}}] = m_{\cdot\cdot\cdot,i} Z_{\cdot\cdot\cdot,i}(\tilde{\Gamma}) \sqrt{2\pi \kappa_{\cdot\cdot\cdot,i}(\psi)} \exp(\psi \sum_{\ell=\tau_{\min}}^{\tau_{\max}} m_{\cdot\cdot\cdot,i} \gamma_{\ell,s,s'} g(x_{s',\cdot,i,t-\ell}))
\]

where \( \mu_{\cdot\cdot\cdot,i} = \mathbb{E}[x_{\cdot\cdot\cdot,i} | x_{\cdot\cdot\cdot,i,t-\tau_{\min}} \ldots x_{\cdot\cdot\cdot,i,t-\tau_{\max}}] \).

### 3.1. PIDC models specifications

This subsection deals with the application of (2) to different statistical settings, associated with binary, count and continuous data. The aim is to assess the versatility of PIDC models to capture different forms of influences induced from well-known statistical families.

**3.1.0.4. Binary data specifications.** Consider data drawn from a multidimensional binary sample space \( \mathcal{X} \in \{0,1\}^{|\mathcal{S}|\mathcal{S}|\mathcal{T}|} \). In this case, we define the underlying measure as \( h(x_{\cdot\cdot\cdot,i}) = 1 \), so that the conditional distribution of individual choices over
item $i$ at time $t$ becomes

$$(x_{i,t} | x_{i,t-\tau_{\min}}, \ldots, x_{i,t-\tau_{\max}}) \sim \text{Bern}(\pi_{i,t}), \quad \text{where } \frac{\pi_{i,t}}{1 - \pi_{i,t}} \propto \exp\left(\sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell}g_{i,t}(t-\ell)\right)$$

Hence, $D(F) = \mathbb{R}^{q}$ as long as $g$ is a well-defined real valued function. This PIDC specification might be seen as an extension to the voter model (see 24 and 39 for more details), where nodes randomly pick at each stage one neighbor and adopt its choice. The binary specification of the PIDC model allows nodes to select choices which have not been selected by their neighbors. Moreover, it allows for non-uniform preferences among neighbors and non-Markovian dependence through the influence structure $\{\Gamma_{\ell}\}_{\ell=\tau_{\min}}^{\tau_{\max}}$.

3.1.0.5. Count data specifications. Consider data drawn from a multidimensional discrete sample space $\mathcal{X} \subseteq \mathbb{Z}^{||I|| \cdot ||S|| \cdot ||T||}$. In this case, we define the underlying measure as $h(x_{sit}) = x_{sit}!$. When no simultaneous dependencies are included ($\tau_{\min} > 0$), individual outcomes have the following distribution:

$$(x_{i,t} | x_{i,t-\tau_{\min}}, \ldots, x_{i,t-\tau_{\max}}) \sim \text{Pois}(\lambda_{it}), \quad \text{where } \lambda_{it} = \exp\left(\sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell}g_{i,t}(t-\ell)\right)$$

Thus, the influence effects result in the joint shift of the conditional mean and variance of each item profile in each time period. Proposition 3.6 establishes a sufficient condition of integrability.

**Proposition 3.6.** In (4) a sufficient conditions for $D(F) = \mathbb{R}^{q}$ is that $g$ is bounded from above.

3.1.0.6. Continuous data specifications. Consider continuous data drawn from a multidimensional continuous sample space $\mathcal{X} \subseteq \mathbb{R}^{||I|| \cdot ||S|| \cdot ||T||}$. In this case, for each $\ell = 1, \ldots, \tau_{\max}$, we assume $\Gamma_{\ell} \in \mathbb{R}^{||S|| \times ||S||}$ be positive definite and define the underlying measure as $h(x_{sit}) = 1/\sqrt{2\pi}$ and $g(x_{sit}) = x_{sit}$. Then we have

$$x_{sit} \ldots x_{i,t-\tau} \sim N(\mu, \Sigma), \quad \text{where } \mu = 0 \quad \text{and } \Sigma = -\frac{1}{2} \tilde{\Gamma}^{-1}.$$  

This reduces to the joint distribution of a vector autoregression of the entire time horizon [49, 59].

8 An application of this continuous data specification in the context of stock price dynamics has been proposed by [52] to study financial contagion. Similarly, [1, 2] derive a static version (simultaneous influences) of (3.1.0.6) from a micro-founded model for a multi-sector economy, where sectors are connected through a supply-chain network which propagates independent Gaussian productivity shocks.

8 An application of this continuous data specification in the context of stock price dynamics has been proposed by [52] to study financial contagion. Similarly, [1, 2] derive a static version (simultaneous influences) of (3.1.0.6) from a micro-founded model for a multi-sector economy, where sectors are connected through a supply-chain network which propagates independent Gaussian productivity shocks.
\( \tau_{\text{min}} > 0 \), by letting \( h(x_{sit}) = 1 \) and \( g(x_{sit}) < 0 \). Then we have

\[
(x_{i,t} \mid x_{i,t-\tau_{\text{min}}} \ldots x_{i,t-\tau_{\text{max}}}) \sim \text{Exp}(\lambda_{it}), \quad \text{where} \quad \lambda_{it} = \sum_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} \Gamma_{i\ell}g_{i\ell,t-\ell}
\]

In this case, influences act in a repulsive way by reducing the probability of individuals to take similar choices.

### 3.2. Further specifications: social media and community structure

In its naive version, the PIDC models establish a conditional distribution of present individual choices as a function of past choices, where the transaction happen through an unknown influence structure that we wish to infer from empirical observations. Sometimes strong assumptions can be made on the pattern of influences: individuals belonging to different groups of homogenous influences, the presence of central nodes which are forced to be more influential than others, etc. These cases are presented in this subsection.

#### 3.2.0.7. Social media dynamics.

Let \( s^* \) be a social media who is able to influence the choices of all individuals in \( S \), without being influenced by anyone. The choices of \( s^* \) on item \( i \) at time \( t \) is denoted as \( x_{s^*it} \). By definition, they are exogenous and cannot be affected by the ones of individuals in \( S \). For every \( s \in S \), the influence of \( s^* \) over \( s \) is included in the kernel of the PIDC model (2) as

\[
\mathbb{P}(x_{sit} \mid x_{i,t-\tau_{\text{min}}} \ldots x_{i,t-\tau_{\text{max}}}) \propto \frac{1}{(h_{ist})^\psi} \exp \left( \psi \sum_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} \left( \sum_{s' \in S} \gamma_{\ell ss'} G_{ss'it\ell} + G_{ss^*it\ell} \right) \right)
\]

where \( G_{ss^*it\ell} = x_{sit}g(x_{s^*it}) \) captures the effect of the choices of \( s^* \) on the ones of \( s \) at time \( t \).

From Proposition 3.4, this implies that the conditional mode in future periods is shifted upward when the selection of item \( i \) from the social media increases. Similar to the models of influencer/imitator dynamics \[62\], in the context of the music industry, this corresponds to the case when the social media broadcasts songs at a high intensity, inducing the reaction of imitators.

For any song \( i \in I \), let \( t_{0i} \) be the starting week when the song is launched. A possible social media specification is to define the exogenous attractiveness trajectory of the \( i^{th} \) song as a Gamma kernel:

\[
g(x_{s^*it}) = \begin{cases} 
\delta_1^0 + \delta_1^1(t - t_{0i}) + \delta_2^2 \log(t - t_{0i}) & \text{if } t > t_{0i} \\
-\infty & \text{otherwise}
\end{cases}
\] (5)
Thus, the social media dynamics is fully characterized by the tuning parameters $\delta^0_i$, $\delta^1_i$, and $\delta^2_i$.

### 3.2.0.8. Community structure.
Consider a set of communities $K_F$ and the function $\kappa_C : \mathcal{S} \rightarrow K_F$ which assigns to each individual $s \in \mathcal{S}$ a given community. Then, for each time lag $\ell$ the influence structure $\Gamma_\ell$ can be restricted as follows:

$$
\gamma_{\ell s_0r} = \gamma_{\ell s_1r}, \quad \text{for all } s_1, s_2, r \in \mathcal{S}, \text{ such that } \kappa_C(s_0) = \kappa_C(s_1) \neq \kappa_C(r),
$$

(6)
i.e., individuals in the same community follow the same influence pattern with respect to other communities. In the context of the music broadcasting industry, the presence of different music formats gives rise to certain form of community structure in the network of broadcasting companies, as discussed in detail in Subsection 4.1.

### 3.2.0.9. Stable influence structure.
Let $\theta_\ell$ be a real quantity capturing the intensity of the influence reaction after $\ell$ periods. Assuming that the topology of the influence structure remains stable across the time lags, the conditional distribution (2) can be rewritten as

$$
\mathbb{P}(x_{sit} | x_{i\cdot t\cdot} \ldots x_{\cdot \cdot t\cdot}) \propto \frac{1}{(h_{int})^\psi} \exp \left( \psi \sum_{\ell=\tau_{\min}}^{\tau_{\max}} \sum_{s' \in \mathcal{S}} \theta_{\ell} \gamma_{ss'} G_{ss'it} \right)
$$

(7)
This model specification allows for a substantial reduction of the dimensionality of the parameter space from $(\tau_{\max} - \tau_{\min}) |\mathcal{S}|^2$ to $(\tau_{\max} - \tau_{\min}) + |\mathcal{S}|^2$.

### 3.3. On the empirical estimation of the influence structure
For each model specification of the PIDC class of models in Definition 3.1, maximum-likelihood can be used to make inference on the influence structure $\hat{\Gamma}$. In fact, this inferential approach relies on well-known properties, which hold for the natural parameters of the exponential family of distributions, such as $\hat{\Gamma}$.

**Proposition 3.7.** Let $\hat{\Gamma}^*$ be the true unknown influence structure to be inferred from the observed panel $\{x_{ist} | i \in \mathcal{I}, s \in \mathcal{S}, t \in \mathcal{T}\}$ and $\hat{\Gamma}^{(\text{MLE})}$ the maximizer of $\mathbb{P}(x_{i\cdot t\cdot} \ldots x_{\cdot \cdot t\cdot} | \hat{\Gamma})$. As a consequence of well-known properties of the maximum likelihood estimation, when the model is identifiable, each of the following statements$^9$ are true:

$^9$Hereby these statements are claimed without a formal proof, as they can be straightforwardly deduced from well known properties of the maximum likelihood estimator, see Theorem 10.1.6 of [18] and Theorem 3.10 of [46].
(i) Since $\gamma_{ij}$ is a natural parameter for the conditional model (1), then the joint distribution (2) is globally concave in $\hat{\Gamma}$.

(ii) $\hat{\Gamma}^{(\text{MLE})}$ converges in probability to $\hat{\Gamma}^*$ when $m$ grows large (consistency).

(iii) The sampling distribution of $\hat{\Gamma}^{(\text{MLE})}$ converges to a Gaussian distribution centered around $\hat{\Gamma}^*$ when $m$ grows (asymptotic normality).

These properties depend on the identifiability of the PIDC class of models, that is, the possibility to learn $\hat{\Gamma}$ after obtaining a finite number of observations from (2). This is equivalent to saying that different influence structures must generate different probability distributions over the choices.

Based on Definition 3.1, identifiability requires that $\mathbb{P}(x_{\tau+1:\tau+m} | \hat{\Gamma}) = \mathbb{P}(x_{\tau+1:\tau+m} | \hat{\Gamma}')$ can only be true if $\hat{\Gamma}' = \hat{\Gamma}$, for all $x \in \mathcal{X}$. Thus, a PIDC specification is non-identifiable if there exist a couple of influence structures $\hat{\Gamma}'$ and $\hat{\Gamma}$, with $\hat{\Gamma}' \neq \hat{\Gamma}$, which give rise to the same probability distribution over the set of dynamic choices $\mathcal{X}$:

$$\frac{1}{Z(\Gamma)} \frac{1}{h_{\Pi}(x)} \exp \left( \psi x^\top \hat{\Gamma} g \right) = \frac{1}{Z(\Gamma')} \frac{1}{h_{\Pi}(x)} \exp \left( \psi x^\top \hat{\Gamma}' g \right), \quad \text{for all } x \in \mathcal{X}. \quad (8)$$

Here we consider a sufficient identification condition, which exclusively relies on the relative sizes of the individual and item dimensions of the panel (properties of $\mathcal{X}$).

Based on the fact that $Z(\hat{\Gamma}) = \sum_x \frac{1}{h_{\Pi}(x)} \exp \left( \psi x^\top \hat{\Gamma} g \right)$, we see that if there exist a couple of parameters $\Gamma'$ and $\Gamma$, with $\Gamma' \neq \Gamma$, such that $x^\top \hat{\Gamma} g = x^\top \hat{\Gamma}' g$, for all $x \in \mathcal{X}$, then for the same couple of parameters it must also be true that $Z(\Gamma) = Z(\Gamma')$. By defining $\bar{\Gamma} = \hat{\Gamma} - \hat{\Gamma}'$, we see that $x^\top \bar{\Gamma} g = 0$ is linear with respect to $\bar{\Gamma}$. Since (8) must be true for all $x \in \mathcal{X}$, this equality can only be verified when $\bar{\Gamma} g = 0$. As a consequence, identifiability is guaranteed as long as $\bar{\Gamma} = 0$ is the unique solution of the linear system $\bar{\Gamma} g = 0$ for all $x \in \mathcal{X}$.

This condition can be expressed in standard matrix form, by reordering the vector of parameters and defining $\bar{\eta} = [\bar{\gamma}_{\tau_{\text{min}}} \ldots \bar{\gamma}_{\tau_{\text{max}}}]^\top$, where $\bar{\gamma}_\ell \in \mathbb{R}^{|S|^2}$ is a vector of differences between influence effects (obtained by stringing out the elements of $\Gamma_{\ell} - \Gamma'_{\ell}$ in lexicographic order), and $\bar{L} = [L_1 \ldots L_{|I|}]^\top$, where
Matrix $\hat{L}$ has $|S||I|(|T| - \tau_{\text{max}} + \tau_{\text{min}})$ rows and $(\tau_{\text{max}} - \tau_{\text{min}})|S|^2$ columns. It turns out that a sufficient condition of identifiability is that the null space of $\hat{L}$ has zero dimension (which implies that the system $\bar{\Gamma}\bar{g} = \hat{L}\eta = 0$ has as a unique solution in the form $\bar{\Gamma} = \hat{\Gamma} - \hat{\Gamma}' = 0$). Thus, the PIDC class of models is identifiable on the subset of $X$ verifying this condition.

In practice, for an observed panel $\{x_{ist} | i \in I, s \in S, t \in T\}$, the uniqueness of the maximum likelihood estimation is guaranteed by the strict concavity of $P(x_{.,.,\tau+1} \ldots x_{.,.,T} | \hat{\Gamma})$ in $\hat{\Gamma}$, as stated in Proposition 3.7. This can most likely happen in a subset of the parameter space. On this respect, it must be noted that, beside providing flexible formulations for real problems, the model specifications presented in Subsection 3.2 play an important role in the identifiability though the constraint structure on the parameter space.

To cast a closer look at their impact on the number of parameters characterizing $\hat{\Gamma}$, we provide an illustrative clarification of the dimensionality of the parameter space. We consider the case of a network of influences forced to verify a community structure, with $c$ groups of equal size, and two time-lag specifications: one unrestricted across the lags (as in the general PIDC formulation in Definition 3.1) and another presenting a fixed network topology across lags (as formulated in (7)). This comparison is illustrated in Figure 3.

The model specification (7) results in a tangible reduction of the number of parameters to be estimated when $n$ and $\tau_{\text{max}}$ become large. In particular, the combination of network partition by communities and fixed influence structure across the lags is capable of reducing the degrees of freedoms up to more than 80% (passing from 30003 parameters to 5007 parameters in the worst case represented in Figure 3).

---

10It can be shown that the probability of all points in $X$ for which the null space of $\hat{L}$ has dimension greater than zero goes to zero when $|I|(|T| - \tau_{\text{max}} + \tau_{\text{min}}) - (\tau_{\text{max}} - \tau_{\text{min}})|S|$ goes to infinity. Intuitively, this means that it becomes highly unlikely to observe linearly dependent sequences of choices when the length of the time window and the number of items involved become large.
(a) \( \tau_{\text{max}} = 1, c = 1 \).
(b) \( \tau_{\text{max}} = 2, c = 1 \).
(c) \( \tau_{\text{max}} = 3, c = 1 \).
(d) \( \tau_{\text{max}} = 1, c = 2 \).
(e) \( \tau_{\text{max}} = 2, c = 2 \).
(f) \( \tau_{\text{max}} = 3, c = 2 \).

Figure 3. Number of parameters to be estimated for different combinations of \( c, n \) and \( \tau_{\text{max}} \). The black bars denote the case of unrestricted influence structure, whose dimensionality is \( \tau_{\text{max}}[c \cdot (n/c)^2 + c^2] \), whereas the red dotted lines denote the case of fixed influence structure across lags, whose dimensionality is \( c \cdot (n/c)^2 + c^2 + \tau_{\text{max}} \).
3.4. Simulation exercise

The consistency and the asymptotic normality of the maximum-likelihood estimation (properties (ii) and (iii) in Proposition 3.7) require an infinite amount of items (which are assumed to be independent in Definition 3.1). Since empirical observations involve a limited amount of items, the analysis of the worst case scenario (i.e., $|I| = 1$) becomes relevant. This subsection explores the sampling distribution of the maximum-likelihood estimation under such a worst case scenario, using different influence structures and model specifications.

As graphically illustrated in the network plots in Figure 4, choices are studied for two topologically different influence structures, a directed star and a directed ring on a society involving 6 individuals (i.e., $|S| = 6$):

\[
\Gamma = \begin{bmatrix}
0 & 1/n & 0 & 0 & 0 & 0 \\
0 & 0 & 1/n & 0 & 0 & 0 \\
0 & 0 & 0 & 1/n & 0 & 0 \\
0 & 0 & 0 & 0 & 1/n & 0 \\
0 & 0 & 0 & 0 & 0 & 1/n \\
1/n & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\quad \text{and} \quad
\Gamma = \begin{bmatrix}
0 & 1/n & 1/n & 1/n & 1/n \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

The computational experiment relies upon a collection of 100,000 simulated panels from the joint distribution (2). For each of the 100,000 replicates, the maximum likelihood estimation of $\Gamma$ has been computed, giving rise to an empirical sampling distribution. To make the estimation more realistic, non-negativity constraints on $\Gamma$ are included, reflecting the bulk of models for social influences [5, 36, 45, 48, 57, 66], where network propagation acts in the sense of lagged similarity or imitation. In such cases, the maximizer of $\mathbb{P}(x_{\cdot,T+1} \ldots x_{\cdot,T} \mid \hat{\Gamma})$ must satisfy side constraints (i.e., $\hat{\Gamma} \geq 0$), resulting in a lower bound of the likelihood value, in comparison with the unrestricted
Table 1. Description of the sampling distribution (over 100,000 simulations from the PIDC models with $\tau_{\text{min}} = \tau_{\text{max}} = 1$) for the pairwise influences parameter associated with zero true values. The same simulation is reproduced for two latent influenced structures (a 6 node ring and a 6 node star) and two time horizons ($T = 20$ and $T = 200$).

<table>
<thead>
<tr>
<th>Family</th>
<th>$T = 20$</th>
<th>$T = 200$</th>
<th>Family</th>
<th>$T = 20$</th>
<th>$T = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ring</td>
<td>Star</td>
<td></td>
<td>Ring</td>
<td>Star</td>
</tr>
<tr>
<td>1st quart</td>
<td>0.00</td>
<td>0.00</td>
<td>1st quart</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Median</td>
<td>0.00</td>
<td>0.00</td>
<td>Median</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.15</td>
<td>0.07</td>
<td>Mean</td>
<td>0.19</td>
<td>0.16</td>
</tr>
<tr>
<td>3rd quart</td>
<td>0.10</td>
<td>0.10</td>
<td>3rd quart</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>True value</td>
<td>0.00</td>
<td>0.00</td>
<td>True value</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The same experiment has been reproduced for two different time windows ($T = 20$ and $T = 200$) and two different statistical families (the conditional Poisson in (4) and the conditional Gaussian in (3.1.0.6)), resulting in 800,000 optimization instances, each involving $n^2 = 36$ decision variables.\(^{11}\)

Tables 1 and 2 report the analysis of the sampling distributions for the pairwise influences parameters. The bottom row of each table reports the true value of the parameters to be estimated, which corresponds to the influence structures depicted in Figure 4.

While this experiment is built upon the worst-case scenario for the PIDC models (the one in which a unique item is observed), the goal here is to assess the inferential properties when the data set is far away from the dimensionality required for the consistency of the maximum-likelihood estimator.

Despite observing a unique item, when influence connections are present in the true structure $\Gamma$, the maximum-likelihood estimation gives rise to higher values than in the cases in which influences are not present. In fact, the estimated PIDC parameters associated with positive values in Figure 4 (i.e., 1/6) are between 0.09 and 0.28 on average, while the the ones associated with zero values are between 0.02 and 0.19 on average (i.e., the topological relations of the Granger causality is seemingly correct). Most importantly, the increase of the observed time window (from $T = 20$ to $T = 200$) reduces the bias even when inference is built on the observation of a unique item. Next to it, tables 1 and 2 reveal a non-negligible role of the probabilistic family and the underlying structures in the convergence result when passing from $T = 20$ to $T = 200$.

\(^{11}\)The optimization of the likelihood function has been curried out using an alternating direction method coded in C++. The algorithm has been initialized at point $\tilde{\Gamma} = 0$ and a tolerance gap of $1.0e - 5$ has been set as a stopping condition. All numerical tests have been performed on a R5500 work-station with processor Intel(R) Xeon(R) CPU E5645 2.40 GHz, and 64 Gbytes of RAM, under a Windows Server 2012 operative system.
Finally, it must be noted that the empirical context presented in the next section relies on a larger panel, comprising a time windows of $T = 255$ weekly observations of $I = 200$ songs in radio stations play-lists.

4. Empirical validation of network discovery in the context of music broadcasting

The statistical properties studied in Section 3 reviled the flexibility and limitations of the network contagion mechanisms described by the PIDC class of models. Based on the different model specifications presented in Subsection 3.1 and Subsection 3.2, hereafter we provide an empirical application, for assessing how the proposed PIDC class of models can be used in practice to detect influences in the context of songs broadcasted on the radio.

4.1. Data on music broadcasting

We collected data about the program schedules of songs played in broadcasting companies (radio stations) in the UK, from January 1, 2011 to March 31, 2015. The data format consists in a sequence of tables identifying a unique moment (day and hour)

<table>
<thead>
<tr>
<th>family = Poisson</th>
<th>1st quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 20$</td>
<td>0.24 0.24 0.24 0.24 0.24</td>
<td>0.28 0.28 0.28 0.28 0.28</td>
<td>0.40 0.41 0.41 0.41 0.41</td>
<td></td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.01 0.01 0.01 0.01 0.00</td>
<td>0.08 0.08 0.08 0.08 0.08</td>
<td>0.17 0.17 0.17 0.17 0.17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>family = Gaussian</th>
<th>1st quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 20$</td>
<td>0.28 0.28 0.28 0.27 0.26</td>
<td>0.27 0.27 0.27 0.27 0.21</td>
<td>0.34 0.34 0.34 0.33 0.33</td>
<td></td>
</tr>
<tr>
<td>$T = 200$</td>
<td>0.08 0.08 0.08 0.08 0.08</td>
<td>0.01 0.01 0.01 0.01 0.01</td>
<td>0.15 0.15 0.15 0.15 0.15</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Description of the sampling distribution (over 100,000 simulations from the PIDC models with $\tau_{\min} = \tau_{\max} = 1$) for the pairwise influences parameter associated with positive true values. The same simulation is reproduced for two latent influenced structures (a 6 node ring and a 6 node star) and two time horizons ($T = 20$ and $T = 200$).
in which a specified song is played in a specified broadcasting company. The database includes information about artists and names of each song, day and time of the day being played, and radio station of that play.\textsuperscript{12}

Table 3 summarizes the sizes of the different dimensions of the data set, covering broadcasting stations over a specified time horizon. The information has been aggregated at a weekly level. The 22 stations are partitioned in accordance with their music format – World-music (2 stations), Contemporary and Easy listening (7 stations), Rock music (6 stations), Top 40 and Urban (7 stations). This classification is based both on the Wikipedia description of the broadcasting companies and the information resulting from their corresponding webpages. Within a given radio format, individual decisions are much more homogeneous and the corresponding number of plays appears to be positively correlated.\textsuperscript{13}

<table>
<thead>
<tr>
<th>Stations</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artists</td>
<td>32,765</td>
</tr>
<tr>
<td>Songs</td>
<td>74,712</td>
</tr>
<tr>
<td>Time periods (weeks)</td>
<td>255</td>
</tr>
</tbody>
</table>

Table 3. Dimensions of panel data set.

Table 4 describes the number of songs and artists played within the analyzed time windows, with different level of share over the total number of plays.

<table>
<thead>
<tr>
<th></th>
<th>50% share</th>
<th>40% share</th>
<th>30% share</th>
<th>20% share</th>
<th>10% share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Played songs</td>
<td>1,180</td>
<td>731</td>
<td>406</td>
<td>190</td>
<td>64</td>
</tr>
<tr>
<td>Played artists</td>
<td>215</td>
<td>135</td>
<td>79</td>
<td>40</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 4. Descriptive information of data set of the UK radio broadcasting industry.

Despite the large amount of songs broadcasted, only few of them are frequently played: 1,180 songs out of 74,712 capture half of the total plays (50% market share). A similar behavior can be observed at the artist level.

\textsuperscript{12}The information was obtained from BMAT, which has developed a technology (called Vericast) to monitor the songs being played in real time on any radio station in any country. (This is done by identifying the musical fingerprint of each song and matching it with their large database of songs.)

\textsuperscript{13}Note that the induced community structure implies $\mathcal{K}_F = \{\text{Contemporary, Top-40, World music, Rock}\}$. When this community structure is not taken into account, the dimensionality of the parameter space of (2) is $(\tau_{\max} - \tau_{\min})|\mathcal{S}|^2$, which grows large (and possibly exceed the one of the sample space) as the number of individuals (stations) increases. The inclusion of this community structure allows reducing the dimensionality, so that the number of influence effects reduces from $22 \times 21 \times (\tau_{\max} - \tau_{\min})$ to $(7 \times 6) + (7 \times 6) + (6 \times 5) + 2) \times (\tau_{\max} - \tau_{\min})$, i.e., a reduction of 75%.
4.2. Estimation

Throughout this subsection, all the estimations rely on the likelihood maximization described in Subsection 3.3 and on the count data specification (4), where we set \( h(x_{sit}) = x_{sit} \). For the comparative analysis two training sets are taken into account. They consist of the top 10 and the top 200 most popular songs.

4.2.0.10. Song life cycle and the influence structure. Firstly, we estimate the dynamic attractiveness of songs after their launch week \( t_0 \) (the exogenous social trend or song life cycle characterized by the parameters \( \{ (\delta_0, \delta_1, \delta_2) \}_{i \in I} \) in (5)), along with the influence structure \( \{ \Gamma_\ell \}_{\ell = \ell_{\min}}^{\ell_{\max}} \). The colored envelopes in figures 5 and 6 show a simulated 95% confidence interval of the number of plays of an arbitrary song, under the estimated models using respectively the top-10 and the top-200 songs as training sets.

In fact, Table 4 revealed that only few songs are quantitatively relevant in the size of the overall broadcasting patterns (the number of songs whose market share is at least a half of the most broadcasted song is just 30).
The estimated social media parameters \( \{(\delta^0_i, \delta^1_i, \delta^2_i)\}_{i \in I} \) reflects the common lifecycle of songs across stations and allows for an aggregate prediction of future dynamics taking into account competing song releases. Although the results are qualitatively similar for 10 and 200 songs, we observe a slightly increased variability when comparing the envelope in Figure 6 (associated to a training set with the first 200 songs), with the one in Figure 5 (associated to a training set with the first 10 songs).

Analogously, based on the same training sets, the network plots in figures 7 and 8 report the estimated influence structures within each of the four radio formats. The sizes of the depicted connections represent the maximum-likelihood estimation of \( \hat{\Gamma} \) and can be interpreted as the sensitivity of the influenced station to increase its number of plays of a song if the influencer station played it more often than the average in previous periods – recall that this average may be driven by the social media dynamic (see specification (5)). When \( \hat{\Gamma} \) is zero, the differentiation between broadcasting companies is controlled by the different effect of the social media \( s^* \) on the corresponding stations.

**4.2.0.11. The time reaction of influence.** The problem of studying a time window in which individuals are influenced (and react to the stimulus of others) is here taken into account from the statistical outlook of the lag-length evaluation [54]. The goal is to determine how fast imitation occurs and to assess whether choices are more or less simultaneous (with weight being put on low lags, e.g., \( \ell = 1 \)) or contagion takes a longer time to propagate (with more weight put on higher lags).

The estimation in Subsection 4.2.0.10 has been carried out using \( \tau_{\text{min}} = \tau_{\text{max}} = 1 \) week. Here we consider a delayed influence with \( \tau_{\text{min}} = 1 \) and \( \tau_{\text{max}} = 10 \) weeks.

Figures 9 summarize in a graphical illustration the distribution of of the estimate \( \{\Gamma_\ell\}_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} \). Each line correspond to the received total influence of each station, whereas the red dotted lines depict the average over each series.

For both training sets, the highest influence seems to appear after one and three weeks. Rock stations are the ones who do imitate each other the most and whose pairwise influences are the last to decline over the time lags. Conversely, contemporary stations and top-40 stations have weaker influences which quickly immediate decay.

A straightforward implication of this level of analysis is the selection of the optimal time lag for the estimation, which is crucial to avoid overfitting in PIDC models. In fact, the high dimensionality of the parameter space of PIDC models entails a parsimonious inclusion of lags in the model specification, after a preliminary estimation of a saturated model. This outlook on model selection is explored in more depth next.

\(^{15}\)The estimated influences between formats are close to zero. The corresponding network plots have been omitted.
Figure 7. Network plots of the estimated pairwise influences between music stations, using the top 10 popular songs as training sets.
Figure 8. Network plots of the estimated pairwise influences between music stations, using the top 200 popular songs as training sets.
4.2.0.12. Model selection. To assess the adequacy and the need of network influences to account for choice variation, a model selection analysis is used, by taking the zero influence model as a null model for station play-list (a model where $\{\Gamma_\ell\}_{\ell=\tau_{\min}}^{\tau_{\max}}$ is fixed at zero, so that the common life-cycle of song across stations $\{(\delta_0^i, \delta_1^i, \delta_2^i)\}_{i \in I}$ is the only driving effect).

Table 5 contains the model comparison based on the log-likelihood functions, the Akaike information criterion and the likelihood ratio, between the null model and three different lags specification of the PIDC class of model.

The tiny p-values of the likelihood ratio tests support the big jump in the model adequacy when passing from reduced model (null model) to larger model ($\tau_{\max} = 1$, $\tau_{\max} = 3$, $\tau_{\max} = 10$), regardless of the strong increase in the number of degrees of freedom (due to the high dimensionality of the influence structure $\{\Gamma_\ell\}_{\ell=\tau_{\min}}^{\tau_{\max}}$). An analogous conclusion can be reached by looking at the Akaike information criterion on the left column of Table 5.

These results on model selection suggest the adequacy of higher-order models ($\tau_{\max} = 10$) in comparison with lower-order ones ($\tau_{\max} = 1$ and $\tau_{\max} = 3$). In particular, and most importantly, the inclusion of the influence structure always results in a significant increase of the likelihood value, despite the high amount of degrees of freedom that it entails.
### Table 5.

Model selection and global quality, based on the Likelihood ratio test and the Akaike information criterion. The first columns on the left report the training set and the model specification. For the Likelihood ratio test, the values into parenthesis correspond to the p-value and the number of degrees of freedom.

<table>
<thead>
<tr>
<th>Training set</th>
<th>Specification</th>
<th>Likelihood ratio</th>
<th>Log-likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top-10 songs</td>
<td>Null model</td>
<td>( \tau_{\text{max}} = 1 )</td>
<td>( &lt;10^{-16}, 127 )</td>
<td>-275881.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau_{\text{max}} = 3 )</td>
<td>( &lt;10^{-16}, 254 )</td>
<td>-152732.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau_{\text{max}} = 10 )</td>
<td>( &lt;10^{-16}, 805 )</td>
<td>-122711.3</td>
</tr>
<tr>
<td>Top-200 songs</td>
<td>Null model</td>
<td>( \tau_{\text{max}} = 1 )</td>
<td>( &lt;10^{-16}, 127 )</td>
<td>-1495377.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau_{\text{max}} = 3 )</td>
<td>( &lt;10^{-16}, 254 )</td>
<td>-1130562.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau_{\text{max}} = 10 )</td>
<td>( &lt;10^{-16}, 805 )</td>
<td>-974115.2</td>
</tr>
</tbody>
</table>

5. Influence maximization under the PIDC class of models

Beyond the ability to predict expected outcomes, well-defined probabilistic models which capture pairwise influences can be helpful in choosing which station is best to diffuse a new song. As mentioned in Section 2, the DeGroot learning model, the linear thresholds model and the independent cascade model are among the most widely studied mathematical framework for diffusion of influence taking place over the edges from node to node. From this outlook, the PIDC class of models can be included in this list.

In this section we examine the influence maximization under the estimated PIDC model and under the assumption of positive influence structure (i.e. influence can only act in the direction of boosting similarities, not dissimilarities).

5.1. Modularity and sensitivity of the influence maximization

Given a full specification of any influence models, an influence maximization problem [32, 42, 55] consists in choosing a good initial set of nodes to target in order to maximize the future diffusion. The expected outcomes of nodes in \( S \) at period \( t_0 + T \) depend on the initial seeds at period \( t_0 \). Let \( S_0 \) be a collection of initial individuals (launching stations who decide to include the new song in their play-list) and \( \mathbf{y}(S_0) \in \{0, 1\}^{|S|} \) the associated indicator vector. The influence of a set \( S_0 \) of nodes at period \( t_0 + T \) is then denoted as \( \sigma(S_0, T) \). Influence maximization problems are usually subject to a fixed cardinality constraint on \( S_0 \), so that the number of initial individuals is fixed. Formally, it can be expressed as the following mathematical program:

\[
\max_{S_0 \subseteq S} \sigma(S_0, T), \quad \text{subject to } |S_0| = k
\]
For most of the influence models, it is NP-hard to determine the optimum set for influence maximization [41]. Moreover, it is not clear how to evaluate $\sigma(S_0, T)$ in polynomial time, as its exact computation requires to integrate over a combinatorial set. [20] and [64] proved that evaluating $\sigma(S_0, T)$ is generally #P-complete both for the linear threshold model and the independent cascade model respectively. However, as shown by [41], it is possible to obtain arbitrarily good approximations in polynomial time, by simulating the random choices and diffusion process. In fact, for each simulated scenario $\kappa$, $\sigma_\kappa(S_0, T)$ is a submodular function both for the linear threshold model and the independent cascade model.

**Definition 5.1 (modularity).** The set function $\sigma(\cdot, T) : 2^{|S|} \rightarrow \mathbb{R}$ is submodular if and only if for all $S_i \subseteq S_j \subseteq S$ and $s \in S$, it verifies

$$\sigma(S_i \cup \{s\}, T) - \sigma(S_i, T) \geq \sigma(S_j \cup \{s\}, T) - \sigma(S_j, T)$$

and it is supermodular if and only if for all $S_i \subseteq S_j \subseteq S$ and $s \in S$, it verifies

$$\sigma(S_i \cup \{s\}, T) - \sigma(S_i, T) \leq \sigma(S_j \cup \{s\}, T) - \sigma(S_j, T).$$

As a consequence of the submodularity result by [41], a natural greedy hill-climbing strategy has been widely adopted to approximate the maximum influence to within a factor of $\left(1 - \frac{1}{e}\right)$, where $e$ is the base of the natural logarithm. Since we focus on a concrete song, we let $|I| = 1$ in what follows, and define the influence function under the PIDC class of model as its expected number of choices at period $T$:

$$\sigma(S_0, T) = \sum_{s \in S} \mathbb{E}[x_{s,t_0+T} | x_{t_0} = y(S_0)] = \int \left\| \frac{1}{Z(\hat{\Gamma}, S_0)} \exp \left( \psi^\top \hat{\Gamma} g(S_0) \right) d\mathbf{z} \right\|_1,$$

where $||\mathbf{z}||_1$ is the one-norm of vector $\mathbf{z}$ and $S_0$ has been included in the normalizing constant notation $Z(\hat{\Gamma}, S_0)$ and in the vector $g(S_0)$ to make the dependency explicit.

For each of the specification of the PIDC conditional model described in Subsection 3.1, closed-form solutions are available for the one-period influence function $\sigma(S_0, 1)$.

**Binary data specification in 3.1.0.4:**

$$\sigma(S_0, 1) = \frac{\exp \left( \Gamma_{\tau_{\text{min}}} y(S_0) \right)}{1 + \exp \left( \Gamma_{\tau_{\text{min}}} y(S_0) \right)}$$

**Count data specification 3.1.0.5:**

$$\sigma(S_0, 1) = \exp \left( \Gamma_{\tau_{\text{min}}} y(S_0) \right)$$

**Continuous data specification 3.1.0.6:**

$$\sigma(S_0, 1) = \Gamma_{\tau_{\text{min}}} y(S_0)$$

---

16The short notation $\sigma(S_0, T)$ refers to the expectation over all scenarios: $\mathbb{E}_n[\sigma_n(S_0, T)]$. 

27
Complementary to the submodularity of $\sigma(S_0, T)$ under the linear threshold model and the independent cascade model, the PIDC class of models gives rise to either supermodular or linear or submodular influence functions, depending on the model specification.

**Proposition 5.2.** Consider the one-period influence function $\sigma(\cdot, 1) : 2^{|S|} \rightarrow \mathbb{R}$ for the PIDC class of model. We claim that:

(i) Under the binary data specification in 3.1.0.4, $\sigma(\cdot, 1)$ is submodular.

(ii) Under the count data specification in 3.1.0.5, $\sigma(\cdot, 1)$ is supermodular.

(iii) Under the continuous data specification in 3.1.0.6, $\sigma(\cdot, 1)$ is linear.

Supermodularity entails increasing marginal returns, so that it becomes worth enlarging the feasible region characterized by $|S_0| = k$, when the optimizer have the possibility to do so. Moreover, as a consequence of the supermodularity result, a natural greedy hill-climbing strategy guarantee the characterization of the optimal initial set $S_0$.

The impact of the initial choices on the expected play-lists strongly depends on the sensitivity of the optimal propagation to the influence structure $\{\Gamma\}_{\ell = \tau_{\max}}^{\tau_{\max}}$.

**Proposition 5.3.** Consider the problem of maximizing the expected number of plays at time $t_0 + T$, under selecting an initial launching station at time $t_0$ and let $S^*$ be the optimal initial set. We claim that

$$\sigma(S_0^*, 1) \geq n |S_0^*| \exp(\psi \rho) \frac{Z(\hat{\Gamma}, S^*)}{Z(\Gamma, S_0^*)},$$

where $\rho$ denotes the spectral radius of $(\Gamma_{\tau_{\min}})^{1/2}$ (this is equivalent to the maximum column sums of absolute values of $(\Gamma_{\tau_{\min}})^{1/2}$).

Thus, the influence matrix acts as an amplifying effect on the initial seed, as the optimal solution is positively related to the spectral radius of the influence matrix.\(^\text{17}\)

The next subsection provides an empirical support in this vein.

### 5.2. On the optimal launching station to maximize songs diffusion

Building on the empirical estimation of the influence structure $\{\Gamma_{\ell}\}_{{\ell = \tau_{\min}}}^{\tau_{\max}}$ using the top-200 songs as a training set (see Subsection 4.2.0.10), the impact of the initial station on future propagation is analyzed hereafter with the aid of the influence maximization theory studied in Subsection 5.1.

\(^{17}\)In particular the lower bound can be conceived as a centrality index in the defined network of pairwise influences [14, 15, 40].
Let us start with the choice of stations. In fact, for each \( s \in S \), the value of the influence function \( \sigma(\{s\}, T) \) provide dynamic scores to all individuals in the network based on their ability to generate positive externalities in the future periods.

Using the Bruno Mars’ songs presented in Section 4.1, tables 6 and 7 report the average values of \( \sigma(\{s\}, T) \) for Locked out of Heaven and Just the way you are respectively, taking each radio station (from the entire collection of UK stations in our data set) as a launching station, where the horizon is \( T = 1, \ldots, 5 \) weeks.

<table>
<thead>
<tr>
<th>Station</th>
<th>Expected plays</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t_0 + 1 )</td>
</tr>
<tr>
<td>106.1 Rock Radio</td>
<td>149.21</td>
</tr>
<tr>
<td>Absolute Radio</td>
<td>146.14</td>
</tr>
<tr>
<td>BBC Radio 1</td>
<td>150.92</td>
</tr>
<tr>
<td>BBC Radio 1 Xtra</td>
<td>142.88</td>
</tr>
<tr>
<td>BBC Radio 2</td>
<td>179.15</td>
</tr>
<tr>
<td>BBC Radio 3</td>
<td>140.24</td>
</tr>
<tr>
<td>BBC Radio 6 Music</td>
<td>187.14</td>
</tr>
<tr>
<td>Capital FM</td>
<td>144.74</td>
</tr>
<tr>
<td>Classic FM</td>
<td>130.24</td>
</tr>
<tr>
<td>Galaxy 102</td>
<td>143.98</td>
</tr>
<tr>
<td>Heart 106.2</td>
<td>143.13</td>
</tr>
<tr>
<td>Kerrang!</td>
<td>141.60</td>
</tr>
<tr>
<td>Key 103</td>
<td>140.29</td>
</tr>
<tr>
<td>Kiss 100 FM</td>
<td>143.58</td>
</tr>
<tr>
<td>Magic 105.4</td>
<td>149.95</td>
</tr>
<tr>
<td>Metro Radio</td>
<td>141.52</td>
</tr>
<tr>
<td>Planet Rock</td>
<td>215.68</td>
</tr>
<tr>
<td>Radio Gold</td>
<td>3502.10</td>
</tr>
<tr>
<td>Smash Hits!</td>
<td>148.54</td>
</tr>
<tr>
<td>Radio City</td>
<td>152.33</td>
</tr>
<tr>
<td>XFM London</td>
<td>160.69</td>
</tr>
<tr>
<td>Ministry Of Sound</td>
<td>142.90</td>
</tr>
<tr>
<td>Observed plays</td>
<td>331</td>
</tr>
</tbody>
</table>

Table 6. Propagation of the broadcasting decision at launch week \( t_0 \) for Locked out of Heaven (Bruno Mars).

Some important insights on diffusion properties of the starting decisions are worth mentioning.

- The main impact of the launching company is propagated one and two weeks after the premiere and reduces after the fourth week for most of the launching stations. A more contextual interpretation is the fact that the long-run behavior of a song is poorly unsensitive to the initial marketing policy.
<table>
<thead>
<tr>
<th>Station</th>
<th>Expected plays</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t_0 + 1$</td>
<td>$t_0 + 2$</td>
<td>$t_0 + 3$</td>
<td>$t_0 + 4$</td>
<td>$t_0 + 5$</td>
</tr>
<tr>
<td>106.1 Rock Radio</td>
<td>148.28</td>
<td>134.59</td>
<td>127.06</td>
<td>122.40</td>
<td>118.31</td>
</tr>
<tr>
<td>Absolute Radio</td>
<td>145.55</td>
<td>131.94</td>
<td>124.72</td>
<td>120.42</td>
<td>117.31</td>
</tr>
<tr>
<td>BBC Radio 1</td>
<td>150.04</td>
<td>135.66</td>
<td>128.03</td>
<td>123.42</td>
<td>120.12</td>
</tr>
<tr>
<td>BBC Radio 1 Xtra</td>
<td>142.44</td>
<td>128.99</td>
<td>121.76</td>
<td>117.62</td>
<td>114.69</td>
</tr>
<tr>
<td>BBC Radio 2</td>
<td>171.73</td>
<td>153.53</td>
<td>143.07</td>
<td>135.46</td>
<td>129.94</td>
</tr>
<tr>
<td>BBC Radio 3</td>
<td>139.97</td>
<td>126.86</td>
<td>119.78</td>
<td>115.63</td>
<td>112.67</td>
</tr>
<tr>
<td>BBC Radio 6 Music</td>
<td>178.34</td>
<td>160.77</td>
<td>147.12</td>
<td>134.96</td>
<td>125.82</td>
</tr>
<tr>
<td>Capital FM</td>
<td>144.18</td>
<td>130.41</td>
<td>122.65</td>
<td>117.86</td>
<td>114.44</td>
</tr>
<tr>
<td>Classic FM</td>
<td>139.97</td>
<td>126.86</td>
<td>119.78</td>
<td>115.63</td>
<td>112.67</td>
</tr>
<tr>
<td>Galaxy 102</td>
<td>143.49</td>
<td>129.92</td>
<td>122.76</td>
<td>118.55</td>
<td>115.53</td>
</tr>
<tr>
<td>Heart 106.2</td>
<td>142.69</td>
<td>129.14</td>
<td>121.87</td>
<td>117.68</td>
<td>114.73</td>
</tr>
<tr>
<td>Kerrang!</td>
<td>141.25</td>
<td>127.06</td>
<td>119.91</td>
<td>115.70</td>
<td>112.74</td>
</tr>
<tr>
<td>Key 103</td>
<td>140.02</td>
<td>126.88</td>
<td>119.79</td>
<td>115.64</td>
<td>112.68</td>
</tr>
<tr>
<td>Kiss 100 FM</td>
<td>143.14</td>
<td>128.80</td>
<td>121.14</td>
<td>116.52</td>
<td>113.35</td>
</tr>
<tr>
<td>Magic 105.4</td>
<td>149.10</td>
<td>134.29</td>
<td>126.17</td>
<td>121.26</td>
<td>117.67</td>
</tr>
<tr>
<td>Metro Radio</td>
<td>141.11</td>
<td>127.76</td>
<td>120.57</td>
<td>116.35</td>
<td>113.33</td>
</tr>
<tr>
<td>Planet Rock</td>
<td>191.90</td>
<td>157.34</td>
<td>136.10</td>
<td>126.23</td>
<td>120.22</td>
</tr>
<tr>
<td>Radio Gold</td>
<td>3326.65</td>
<td>2853.51</td>
<td>1235.92</td>
<td>818.34</td>
<td>876.53</td>
</tr>
<tr>
<td>Smash Hits!</td>
<td>147.76</td>
<td>133.40</td>
<td>125.68</td>
<td>121.10</td>
<td>118.05</td>
</tr>
<tr>
<td>Radio City</td>
<td>151.09</td>
<td>132.48</td>
<td>123.24</td>
<td>117.93</td>
<td>114.35</td>
</tr>
<tr>
<td>XFM London</td>
<td>226.11</td>
<td>202.94</td>
<td>292.60</td>
<td>179.29</td>
<td>170.87</td>
</tr>
<tr>
<td>Ministry Of Sound</td>
<td>142.22</td>
<td>128.25</td>
<td>121.04</td>
<td>116.86</td>
<td>114.06</td>
</tr>
</tbody>
</table>

Table 7. Propagation of the broadcasting decision at launch week $t_0$ for *Just the way you are* (Bruno Mars).

- The above comments is not valid for Radio Gold\textsuperscript{18}, whose initial diffusion is able to induce an average airplay (the frequently a song is being played on radio stations) of more than 20 daily plays per station in the first week for both *Locked out of Heaven* and for *Just the way you are*. An average airplay of more than 8 daily plays per station still appear after 5 weeks.
- Classic FM\textsuperscript{19} is the station with the smaller propagation effect.
- The influence happening between stations of different formats are very small in size, suggesting that the main core of imitation happen within stations of the same music formats.

\textsuperscript{18}Radio Gold is a network of stations broadcasting in the entire UK, formed by the merger of the Capital Gold Network and the Classic Gold Network in August 2007. In our data set, it is associated to *Contemporary and Easy Listening*.

\textsuperscript{19}Classic FM is an Independent National Radio stations broadcasting classical music in the entire UK since 1992. In our data set, it is associated to *World Music*. 

30
These insights deserve a careful interpretation. As already mentioned in Section 1, statistical models fail to distinguish between causality and co-variation, so that the only evidence detected by the estimated model relies on the Granger causality. As a result, influence is invoked when sequences of individual decisions are systematically anticipated by the ones of other individuals, as is the case for Radio Gold in the aforementioned results. Thus, under the inductive belief, such an observed pattern of anticipations can be propagated to obtain the expected diffusion in tables 6 and 7.

6. Conclusions

In this paper, we developed a twofold modeling framework, jointly supporting the influence discovery (based on a large variety of specifications) and the influence maximization (building on the modularity conditions of the different specifications). As shown in Section 3.1, the generality of our approach comes from the fact that many statistical models belonging to the exponential family can be seen as specific cases of the PIDC class of models, sharing common properties that we have uncovered in Section 3 (for the influence discovery part) and Section 5.1 (for the influence maximization part). Thus, the core idea of lead-lag synchronization in multiple choices (empirically motivated by the observation of songs diffusion on the radio) has resulted in a twofold modeling framework, integrating a variety of restricted formulations, which have been so far considered in isolation.

Contextually, the core contributions and results of this work must be highlighted as follows:

- On the modeling side, mirroring the dynamics of songs diffusion, we propose a parametric approach for two purposes (which are traditionally studied in separate streams of literature): (i) the one of influence discovery, by which observed choices are tracked to infer patterns of pairwise influences, and (ii) the one of influence maximization, by which the initial choices of individuals can be used to optimize the contagion on the overall network.
- On the probabilistic side, we uncover a number of statistical properties of the proposed class of models, to analyze the conditional moments, the parameters sensitivity and the identifiability conditions.
- On the empirical side, the PIDC class of models was crafted for fashion goods which experience significant dynamic variations in popularity, and was applied to the diffusion of songs over time. Our results provided an empirical estimation of the strength of station-to-station influences, which can be used as a score for the ability of conditioning the future choices of other players in the market. We found evidence that there were influencing links (one station copying another).
Finally, building on the influence maximization part of the PIDC class of models and with knowledge of the influence structure, we were able to support the decision of music producers about which station to choose for first diffusing their products. We showed that influence effects between broadcasting companies allows maximizing the diffusion of a song in the first weeks after launch, but these effects fade away with time as the song reaches the entire network.

The proposed methodology is general and might be extended to many other multidimensional panel concerning dynamic choices over multiple items. Applying the collection of techniques developed in this paper to other social and economic settings may generate further insights on how to discover influence relationships and use them to generate propagation effects in the network.

References


[52] S. Nasini and D. Erdemlioglu, Multiple channels of financial contagion: an empirical
Appendix A: Proofs

0.0.13. **Proposition 3.2.**

**Proof.** Let $z$ be a random vector (with dimension $d_z$) defined on a suitable sample space $Z$, $h : Z \rightarrow \mathbb{R}$ be non-negative and non-decreasing, and consider an exponential random model with vector of natural parameter and sufficient statistics $\gamma$ (with dimension $b$) and $T(z)$ respectively. We defined $q_\gamma(z) = \exp(T(z)\gamma)/h(z)$. The expectation
of \( z \) can be written as:

\[
E[z] = \frac{1}{Z(\gamma)} \int_{z \in \mathbb{Z}} z q_{\gamma}(z) \, dz
\]

Differentiating with respect to \( \gamma \) we have

\[
\frac{\partial}{\partial \gamma} E[z] = \left( \frac{1}{Z(\gamma)} \right) \frac{\partial}{\partial \gamma} \left( \int_{z \in \mathbb{Z}} z q_{\gamma}(z) \, dz \right) + \left( \int_{z \in \mathbb{Z}} z q_{\gamma}(z) \, dz \right) \frac{\partial}{\partial \gamma} \left( \frac{1}{Z(\gamma)} \right)
\]

where \( \frac{\partial}{\partial \gamma} \left( \frac{1}{Z(\gamma)} \right) \) is a row vector with dimension \( d_\gamma \), and \( \int_{\mathbb{Z}} z q_{\gamma}(z) \, dz \) is a column vector with dimension \( d_z \). Thus, \( \left( \int_{\mathbb{Z}} z q_{\gamma}(z) \, dz \right) \frac{\partial}{\partial \gamma} \left( \frac{1}{Z(\gamma)} \right) \) is a \( d_z \times d_\gamma \) matrix. Similarly

\[
\frac{\partial}{\partial \gamma} \left( \int_{\mathbb{Z}} z q_{\gamma}(z) \, dz \right) = \int_{\mathbb{Z}} z \frac{\partial}{\partial q_{\gamma}} q_{\gamma}(z) \, dz
\]

is a \( d_z \times d_\gamma \) matrix. Note that the two components of \( \frac{\partial}{\partial \gamma} E[z] \) correspond to:

\[
\left( \frac{1}{Z(\gamma)} \right) \frac{\partial}{\partial \gamma} \left( \int_{\mathbb{Z}} z q_{\gamma}(z) \, dz \right) = \left( \frac{1}{Z(\gamma)} \right) \left( \int_{\mathbb{Z}} z \frac{\partial}{\partial \gamma} q_{\gamma}(z) \, dz \right) = \mathbb{E} [z \, T(z)^\top]
\]

\[
\left( \int_{\mathbb{Z}} z q_{\gamma}(z) \right) \frac{\partial}{\partial \gamma} \left( \frac{1}{Z(\gamma)} \right) = \left( \int_{\mathbb{Z}} z q_{\gamma}(z) \right) \frac{\partial}{\partial \gamma} \left( \frac{1}{Z(\gamma)} \right)
\]

\[
= \left( \int_{\mathbb{Z}} z q_{\gamma}(z) \right) \frac{\partial}{\partial \gamma} \left( -\frac{1}{Z(\gamma)^2} \right) \left( \int_{\mathbb{Z}} T(z)^\top q_{\gamma}(z) \, dz \right)
\]

\[
= -\mathbb{E} [z] \, \mathbb{E} [T(z)^\top]
\]

Thus, \( \frac{\partial}{\partial \gamma} E[z] = \mathbb{E} [z \, T(z)^\top] - \mathbb{E} [z] \, \mathbb{E} [T(z)^\top] \), which is a \( d_z \times d_\gamma \) matrix containing the covariances between each pair of elements of vectors \( z \) and \( T(z) \). In the case of the the conditional model (1), the sufficient statistic is \( T_{\ellss'} = x_{s,i,t}g(x_{s'i,t-\ell}) \). Thus, we claim that

\[
\frac{\partial}{\partial \gamma_\ellss'} \mathbb{E} [x_{sit} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}}] = g(x_{s'i,t-\ell})(\mathbb{E} [(x_{s,i,t-\ell})^2 \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}}]
\]

\[
-\mathbb{E} [x_{s,i,t} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}}]^2
\]

\[
= g(x_{s'i,t-\ell})\mathbb{V} [x_{sit} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}}],
\]

(1)
for each $i \in I$, $s \in S$. \hfill \Box

0.0.14. Proposition 3.3.

Proof. We start by noting that the unimodality of $P(x_{i,t} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}})$ is equivalent to the unimodality of its logarithm. So we define $f : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$f(x_{i,t}) = \log P(x_{i,t} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}})$$

and $H : \mathbb{R}^{|S|} \rightarrow \mathbb{R}^{|S|}$ as $H(x_1, \ldots, x_n) = [h'(x_1)/h(x_1) \ldots h'(x_n)/h(x_n)]^T$. A sufficient condition for $f$ to be unimodal is that, for any positive $\varepsilon$, it verifies the following two implications:

i) if $\nabla f(z) < 0$ then $\nabla f(z + \varepsilon) < 0$ (componentwise);

ii) if $\nabla f(z) > 0$ then $\nabla f(z - \varepsilon) > 0$ (componentwise).

Note that $\nabla f(x_{i,t}) = \sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell} g_{s,i,t-\ell} - H(x_{i,t})$. Since $\delta = \sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell} g_{s,i,t-\ell}$ is a constant, based on Assumption 1, for any component $s \in S$ we can verify that

i) if $\delta_s < H_s(x_{i,t})$ then $\delta_s < H_s(x_{i,t} + \varepsilon)$;

ii) if $\delta_s > H_s(x_{i,t})$ then $\delta_s > H_s(x_{i,t} - \varepsilon)$;

where $\delta_s$ and $H_s$ are the $s^{th}$ components of $\delta$ and $H$ respectively. \hfill \Box

0.0.15. Proposition 3.4.

Proof. Consider the conditional distribution (1) and maximize it with respect to $x_{i,t}$. From Proposition 3.3 we define $\nabla f(x_{i,t})$ and obtain the first order condition for the mode of $x_{i,t} \mid x_{i,t-\tau_{\min}} \ldots x_{i,t-\tau_{\max}}$

$$\sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell} g_{s,i,t-\ell} = \left[ \begin{array}{c} \frac{h'(\text{mod}(x_{1,t}))}{h(\text{mod}(x_{1,t}))} \\ \vdots \\ \frac{h'(\text{mod}(x_{n,t}))}{h(\text{mod}(x_{n,t}))} \end{array} \right], \text{ and } \text{mod}(x_{i,t}) = H^{-1}\left(\sum_{\ell=\tau_{\min}}^{\tau_{\max}} \Gamma_{\ell} g_{s,i,t-\ell}\right)$$

The second equality comes from the fact that Assumption 1 guarantees that $H$ is invertible. \hfill \Box

0.0.16. Proposition 3.5.

Proof. Let $w : \mathbb{R} \rightarrow \mathbb{R}$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ and consider the Laplace method used to
approximate integrals of the form

\[ \int \limits_{\mathcal{Y}} w(y) \exp(\psi f(y)) dy \approx \sqrt{\frac{2\pi}{|f''(\text{mod}(y))|}} w(\text{mod}(y)) \exp(\psi f(\text{mod}(y))), \quad \text{as } \psi \text{ grows large,} \]

where \( \text{mod}(y) \in \mathcal{Y} \) is the maximizer of \( f(y) \) and \( |f''(\text{mod}(y))| \) is the absolute value of the second derivative of \( f \) at point \( \text{mod}(y) \). In our case, if our aim is to calculate \( \mathbb{E}[x_{sit} \mid x_{i,t-\tau_{\text{min}}} \cdots x_{i,t-\tau_{\text{max}}}] \), based on the conditional PIDC distribution (1), we define

\[
\begin{align*}
f(x_{sit}) &= \psi \left( \sum_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} x_{sit} \gamma_{\ell,s,s'} g(x_{s',i,t-\ell}) - \log h(x_{sit}) \right) \quad \text{and} \quad w(x_{.,it}) = x_{sit}.\end{align*}
\]

Based on Proposition 3.4, we define \( \mathbf{m}_{.,it} = \text{mod}[x_{.,it}] = [m_{1,it} \cdots m_{n,it}]^\top \) and note that

\[
\begin{align*}
\frac{\partial^2}{\partial x_{sit} \partial x_{sit}} f(x_{sit}) \bigg|_{x_{sit}=m_{sit}} &= \psi \left( \frac{h''(m_{sit})}{h(m_{sit})} - \left( \frac{h'(m_{sit})}{h(m_{sit})} \right)^2 \right),
\end{align*}
\]

By letting \( \kappa_{sit}(\psi) = \psi h(m_{sit}) \psi \left( \frac{h''(m_{sit})}{h(m_{sit})} - \left( \frac{h'(m_{sit})}{h(m_{sit})} \right)^2 \right) \) and replacing the values of \( m_{sit}, w, f \) and \( f'' \) corresponding to the PIDC model (1) into the Laplace functional form for integral approximation, we have

\[
\begin{align*}
\mathbb{E}[x_{sit} \mid x_{i,t-\tau_{\text{min}}} \cdots x_{i,t-\tau_{\text{max}}}] &= \frac{m_{sit}}{Z_{sit}(\Gamma)} \sqrt{\frac{2\pi}{h''(m_{sit}) h(m_{sit}) - \left( \frac{h'(m_{sit})}{h(m_{sit})} \right)^2}} \exp(\psi f(m_{sit})) \\
&= \frac{m_{sit}}{Z_{sit}(\Gamma)} \sqrt{\frac{2\pi}{\kappa_{sit}(\psi)}} \exp(\psi \sum_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} m_{sit} \gamma_{\ell,s,s'} g(x_{s',i,t-\ell}))
\end{align*}
\]

.0.0.17. Proposition 3.6.

Proof. For every item \( i \in \mathcal{I} \) and every time \( t \in \mathcal{T} \), consider model (2) and the probability of the total individual outcomes:

\[
P \left( \sum_{s \in S} x_{sit} = y_{it} \mid x_{i,t-\tau_{\text{min}}}, \ldots, x_{i,t-\tau_{\text{max}}} \right) \propto Q_{\gamma}(y_{it}) := \sum_{x_{s,t} \cdots x_{s',t}=y_{st}} \frac{1}{h_{sit}} \exp \left( \psi \sum_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}} \gamma_{s,s'} G_{ss't} \right)
\]

Let \( \gamma_{\text{max}} \) be the maximum element among the influence structure \{\Gamma_{\ell}\}_{\ell=\tau_{\text{min}}}^{\tau_{\text{max}}}.\) Then the
following upper limit can be deduced:

$$Q_{\gamma}(y_{st}) \leq \exp\left( (\tau_{\text{max}} - \tau_{\text{min}}) \gamma_{\text{max}} (y_{st} g(y_{st})) \right) \sum_{x_{s1} + \ldots + x_{|[S]|} = y_{st}} \prod_{r \in S} \frac{1}{x_{sr}!} \max\{\psi, 1\}$$

$$= \left[ \frac{|S|^{y_{st}}}{y_{st}!} \exp\left( \gamma_{\text{max}} (y_{st} g(y_{st})) \right) \right]^{\max\{\psi, 1\}} : \text{[multinomial theorem]}$$

Since $Q_{\gamma}(y_{st}) < \infty$ for any real $y_{st}$, the a normalizing constant exists when $Q_{\gamma}(y_{st})$ goes to zero when $y_{st}$ grows large. By applying the ratio test to the convergence of the series, we find

$$L = \lim_{y \to \infty} \frac{|S|^{y+1} y! \exp\left( \gamma_{\text{max}} ((y + 1) g(y + 1)) \right)}{|S|^y (y + 1)! \exp\left( \gamma_{\text{max}} (y g(y)) \right)}$$

$$= \lim_{y \to \infty} \frac{|S|^y}{y} \exp\left( \gamma_{\text{max}} ((y + 1) g(y + 1)) - (y g(y)) \right)$$

Note that a sufficient condition for $Z(\tilde{\Gamma}) < \infty$ for all $\tilde{\Gamma}$ is that $L < 1$. Convergence is guaranteed if $g$ is non-decreasing and bounded from above, then, for all $y \geq 0$, we have $(y + 1) g(y + 1) - y g(y) \leq 1$.  

\[ \square \]

.0.0.18. Proposition 5.2.

\textbf{Proof.} To keep notation short let $p_i = \Gamma_{\tau_{\text{min}}, y(S_i)}$ and consider the respective condition of submodularity, supermodularity and linearity for the one period influence function $\sigma(., 1) : 2^{|S|} \longrightarrow \mathbb{R}$:

$$\sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) \geq \sigma(S_j \cup \{s\}, 1) - \sigma(S_j, 1), \quad \text{for all } S_i \subseteq S_j \subseteq S \text{ and } s \in S.$$  

$$\sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) \leq \sigma(S_j \cup \{s\}, 1) - \sigma(S_j, 1), \quad \text{for all } S_i \subseteq S_j \subseteq S \text{ and } s \in S.$$  

$$\sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) = \sigma(S_j \cup \{s\}, 1) - \sigma(S_j, 1), \quad \text{for all } S_i \subseteq S_j \subseteq S \text{ and } s \in S.$$  

To prove (i) let us consider the binary data specification (3.1.0.4) and write

$$\sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) = \frac{\exp(p_i + p_s)}{1 + \exp(p_i + p_s)} - \frac{\exp(p_i)}{1 + \exp(p_i)}$$

$$= \frac{\exp(p_i) (\exp(p_s) - 1)}{(1 + \exp(p_i)) (1 + \exp(p_i + p_s))}$$

39
From the submodularity condition the following inequalities are equivalent:

\[ \frac{\exp(p_i)}{(1 + \exp(p_i))(1 + \exp(p_i + p_s))} \geq \frac{\exp(p_j)}{(1 + \exp(p_j))(1 + \exp(p_j + p_s))} \]

\[ \exp(p_i)(1 + \exp(p_i))(1 + \exp(p_i + p_s)) \geq \exp(p_j)(1 + \exp(p_j))(1 + \exp(p_j + p_s)) \]

\[ \exp(p_i) + \exp(2p_j + p_i + p_s) \geq \exp(p_j) + \exp(2p_i + p_j + p_s) \]

\[ \exp(p_i)(1 - \exp(p_i + p_j + p_s)) \geq \exp(p_j)(1 - \exp(p_i + p_j + p_s)) \]

\[ \exp(p_i) \leq \exp(p_j) \]

Since \( \exp(p_i) \leq \exp(p_j) \) is true by the definition of \( S_i \subseteq S_j \subseteq S \) and \( s \in S \), we conclude that under the binary data specification (3.1.0.4), the submodularity condition (.0.0.18) is verified.

To prove (ii) let us consider the count data specification (4) and write

\[ \sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) = \exp(p_i)(\exp(p_s) - 1) \]

From the supermodularity condition the following inequalities are equivalent:

\[ \exp(p_i)(\exp(p_s) - 1) \leq \exp(p_j)(\exp(p_s) - 1) \]

\[ \exp(p_i) \leq \exp(p_j) \]

Since \( \exp(p_i) \leq \exp(p_j) \) is true by the definition of \( S_i \subseteq S_j \subseteq S \) and \( s \in S \), we conclude that under the count data specification (4), the submodularity condition (.0.0.18) is verified.

Finally, to prove (iii) let us consider the continuous data specification (3.1.0.6) and write

\[ \sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) = p_i + p_s - p_i \]

From the linearity condition we have \( \sigma(S_i \cup \{s\}, 1) - \sigma(S_i, 1) = p_s = \sigma(S_i \cup \{s\}, 1) \), for all \( S_i \subseteq S_j \subseteq S \) and \( s \in S \). Thus, under the continuous data specification (3.1.0.6), the linearity condition (.0.0.18) is verified.

.0.0.19. Proposition 5.3.

Proof. Consider the influence function (5.1) and let \( S_0^* \) be the optimal first stage solution of the influence maximization problem (9), under a given PIDC model specification. Note that for any \( S_j \in 2^{S} \), verifying the capacity constraint, we have
\( \sigma(S_0^*, 1) \geq \sigma(S_j, 1) \). We see that

\[
\sigma(S_0^*, 1) \geq \frac{n}{Z(\Gamma, S_0^*)} \frac{\exp \left( \psi u^\top \Gamma_{\tau_{\min}} u \right)}{h_\Pi(u)},
\]

where \( u \) is an arbitrary vector with norm \( |S_0^*| \). Since this is true for all \( u \), it is also true for the maximizer of \( u^\top \Gamma_{\tau_{\min}} u \), subject to \( ||u|| = |S_0^*| \), which is known to be the norm of \( (\Gamma_{\tau_{\min}})^{1/2} \). Thus, by using Assumption 1, we can write

\[
\sigma(S_0^*, 1) \geq \frac{n |S_0^*|}{Z(\Gamma, S_0^*)} \exp \left( \psi \left( (\Gamma_{\tau_{\min}})^{1/2} \right) \right) \geq n |S_0^*| \frac{\exp(\psi \rho)}{Z(\Gamma, S_0^*)}
\]

where \( \rho \) is the spectral radius of \( (\Gamma_{\tau_{\min}})^{1/2} \). \( \square \)