

Modeling Contagious Merger and Acquisition via Point Processes with a Profile Regression Prior

Junchi Yan^{13*}, Shuai Xiao², Changsheng Li³, Bo Jin¹
 Xiangfeng Wang¹, Bin Ke⁴, Xiaokang Yang², Hongyuan Zha¹

¹East China Normal University, ²Shanghai Jiao Tong University

³IBM Research - China, ⁴National University of Singapore

{jcyang,bjin,xfwang,zha}@sei.ecnu.edu.cn, {benjaminforever,xkyang}@sjtu.edu.cn
 lcscheng@cn.ibm.com, bizk@nus.edu.sg

Abstract

Merger and Acquisition (M&A) has been a critical practice about corporate restructuring. Previous studies are mostly devoted to evaluating the suitability of M&A between a pair of investor and target company, or a target company for its propensity of being acquired. This paper focuses on the dual problem of predicting an investor’s prospective M&A based on its activities and firmographics. We propose to use a mutually-exciting point process with a regression prior to quantify the investor’s M&A behavior. Our model is motivated by the so-called contagious ‘wave-like’ M&A phenomenon, which has been well-recognized by the economics and management communities. A tailored model learning algorithm is devised that incorporates both static profile covariates and past M&A activities. Results on CrunchBase suggest the superiority of our model. The collected dataset and code will be released together with the paper.

1 Introduction

Predicting investor’s M&A Merger and Acquisition (M&A) has become a popular business practice¹, allowing firms to instantly acquire new competencies by different dimensions such as brand, channel, technology. M&A has become a major vehicle for company growth since the 1980s [Very *et al.*, 2012]. They can help optimize the market structure and increase market power, obtain tax advantage, generate economies of scale and other synergies, or serve managerial ambitions [Gugler and Konrad, 2002].

Predicting an investor’s prospective M&A *in an arbitrary future time point/period* is challenging and of particular interest to both company executives and market institutions. Anticipating mergers and acquisitions (M&A) helps executives and investors to design their firms strategies and decide

*Corresponding author. The work is supported by STCSM (15JC1401700, 14XD1402100), NIH/NSF R01 GM108341, NSFC (61527804, 61221001, 61129001), the 111 Program (B07022). The first two authors contribute equally to this work.

¹The distinction between ‘merger’ and ‘acquisition’ has become blurred. The paper uses *M&A* and *acquisition* interchangeably.

on their investments. Variations in M&A activity have been found to influence the value created by acquirers. For instance, M&A activity is characterized by periods of waves [Resende, 2008] and acquirers market value statistically increases more significantly for the early movers in a wave. This is why M&A activity forecasts are especially important for many economic stakeholders: prospective M&As help national and international institutions to anticipate Foreign Direct Investment flows, investors can track periods of intense M&A activity in order to acquire shares in future M&A targets. Executives can anticipate movements in the industry.

Limitation of classification methods Despite its imperative business value and scientific interest, modeling and predicting M&A for an *individual* investor is rarely studied. [Pasiouras and Gaganis, 2007] build supervised binary classification models by employing static profiles, and some aggregated statistical indicators about past activities e.g. total number of events, frequency as features for M&A prediction.

Such classification approaches simplify the problem and suffer limitations: they truncate the observation window to an ad-hoc period which induces the label for training set. In fact they are inherently not a behavioral model, and unable to flexibly capture the dynamics of past M&A events nor the prospective events in continuous time space.

The contagious wave-like M&A phenomenon Different from classification/regression models, our point process based method is tailored to a long-standing and well-known phenomenon (to corporate finance and management) – M&A activities tend to spread across both intra-industry and inter-industry *contagiously* [Öberg and Holtström, 2006] forced by various spontaneous and extraneous drivers, e.g. see [Mitchell and Mulherin, 1996; Mariana, 2012], which in macro exhibits the temporal clustering pattern [Maksimovic *et al.*, 2013] or in a more expressive term: ‘waves’ [Brealey, 2012]. For instance, in [Öberg and Holtström, 2006], the authors empirically confirmed that ‘*Following an initial M&A, the M&A pendulum swings back and forth between the customers and suppliers*’. In fact, since the 1890s there are 6 major M&A waves. Our initial inspection to the acquisition records from CrunchBase also verifies this hypothesis as illustrated in Fig.1. Such a contagious nature of M&A activities, to some extent, analogously relates to the viral diffusion in epidemiology, social science and many other disciplines.

Key idea and contributions To model the above contagious

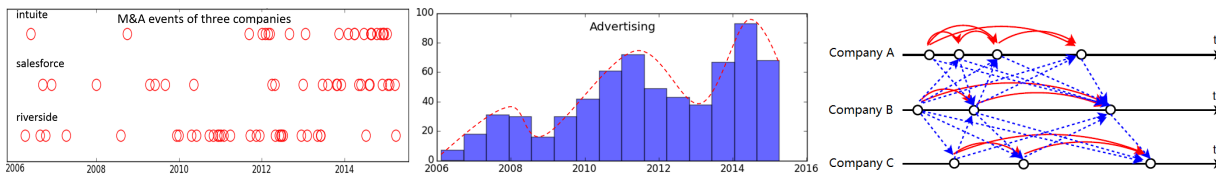


Figure 1: Wave-like contagious M&A exists in both company and industry level. It can be modeled by a Hawkes process.

and wave-like phenomenon, we utilize Hawkes processes [Hawkes, 1971], a special class of point processes, whose intensity function implies how likely an event will happen at each timeframe. The intensity function of Hawkes comprises of a spontaneous intensity and a positive influence of the past events on the current one. Such a positive influence is originated from the self/mutually-exciting property that the occurrence of a past event raises the probability of events happening in future. We find that the Hawkes’s self/mutually-exciting property coincides with the fact that investors’ M&A influences each other, and occur in a cascading fashion.

There are three contributions of our work:

i) To our best knowledge, it is the first time to adopt a point process framework for M&A events modeling and prediction, and the profile covariates is incorporated by a regression prior on the background rate of the intensity function.

Specifically, it is the first time, to the best of our knowledge, for adapting a mutually-exciting point process to model the contagious ‘wave-like’ phenomenon of M&A, though this phenomenon has been a well-established stylized fact to economist/management community for a long time.

ii) We propose a learning algorithm for our regression prior based point process model by Lagrangian relaxation.

iii) We verify our model on a recent real-world dataset of CrunchBase, which we believe is of particular interest to both policy-makers and market players, especially considering we are in the era that the high-tech industry is booming, and disruptive technologies and new business models are emerging.

2 Related Work

Empirical study News medium often issue M&A activity predictions: consulting firms (McKinsey, PWC, or KPMG), investment banks (Goldman Sacks, Morgan Stanley), and institutions regularly put forward their own forecasts about M&A numbers, the value of the M&A market, or foreign direct investments through acquisitions. This is often done by interviewing a few experts or top executives who give their opinion about the future [Öberg and Holtström, 2006].

Algorithmic methods and problem categories By the application scenario, existing algorithmic methods can be categorized into three contexts: i) [Hamilton, 1989; Resende, 2008] forecast the M&A at *aggregated country/industry level* based on macroeconomic variables e.g. GDP, money supply, stock market growth. A predictive model such as Kalman filter is applied [Very *et al.*, 2012]; ii) measure the suitability between a *specific bidder company and a target one* [Weber and Dholakia, 2000; Gugler and Konrad, 2002; Song and Chu, 2006; Pasiouras and Gaganis, 2007], whereby financial and managerial variables are often exploited by a classification model;

iii) estimate the possibility of a target company being acquired [Slowinski *et al.*, 1997; Ragothaman *et al.*, 2003; Ali-Yrkkö *et al.*, 2005; Pasiouras and Gaganis, 2007; Xiang *et al.*, 2012]. The business motivation is that a target company with higher chance of being acquired is often more valuable to potential investors. Logistic Regression is widely adopted to build the prediction model [Ali-Yrkkö *et al.*, 2005; Pasiouras and Gaganis, 2007].

Predicting investor’s M&A Compared with the above three categories of prediction problems, directly predicting individual investors’ M&A event is less studied. There are few methods [Ragothaman *et al.*, 2003] addressing this problem while they use Logistic regression similar to the target company classification problem [Xiang *et al.*, 2012]. In contrast, point process formulation can naturally incorporate M&A’s timestamp. The fact that M&As are often in the wave forms which motivates to design a tailored point process model.

Hawkes processes This point process is originated from [Hawkes, 1971; Hawkes and Oakes, 1974]. It is a one-dimensional self-exciting point process modeling the event sequences that exhibit temporal clustering patterns over time. Its main characteristics is the modeling of chain relation effects, the occurrence of one event triggers that of another. The early application of Hawkes process model [Ogata, 1988; 1998] refers to model the occurrences of earthquakes.

There are also multi-dimensional i.e. mutually-exciting Hawkes process variants where the triggering effect of different types of events are considered. The mutually-exciting Hawkes process has been widely used to model social behaviors on networks [Blundell *et al.*, 2012; Zhou *et al.*, 2013a]. There is no work on modeling M&A by Hawkes processes, though the wave-like phenomenon for M&A can be a good testbed for extending the boundary of Hakes processes.

Contagious wave-like M&A Starting with Nelson [Nelson, 1959], a long line of empirical research has shown that merger activities tend to cluster in time and the events cascade contagiously. [Harford, 2005; Resende, 2008] study the underlying drivers for the ‘wave’ phenomenon and reach a consensus about its existence across different countries and industries. There are two main explanations by these works: the neoclassical hypothesis ascribes the clustering waves to technology or regulatory shocks [Jovanovic and Rousseau, 2001; 2002]. The other behavioral hypothesis [Shleifer and Vishny, 2003] posits the temporal clustering of M&A activity is driven by stock market valuations. Bull markets lead groups of bidders with overvalued stock to use it to buy real assets of undervalued targets through M&A. The competition between the bidders further exaggerates the temporal clustering phenomenon of M&A waves. They also find company who is undertaking strategy transition, tend to repeatedly conduct

M&A in a short time window. [Hou *et al.*, 2015] take the M&A activity from a complex network perspective, and state that the cascading-failure phenomenon is due to the breakdown of one or more nodes or edges may lead to the failure of others via their coupling mechanisms.

CrunchBase It has become a popular data source as it maintains abundant information about M&A and investments mostly for North America (<https://www.crunchbase.com/>). [Eugene and Yuan, 2012] perform prescriptive data mining on the CrunchBase and uncover several general rules for companies seeking investment. The similar acquisition target identification problem is studied in [Wei *et al.*, 2009], where technological variables derived from patent analysis and profiles of investor and candidate target company are used for an ensemble classification model. [Xiang *et al.*, 2012] find that the factual and topic features using profiles and news articles on CrunchBase are also indicative and more readily accessible than the social relation data used in [Eugene and Yuan, 2012; 2013]. They define the M&A prediction problem by classifying the candidate companies into M&A target and non-targets, rather than a point process method and irrelevant to the investor. On the contrary, this paper predicts a specific investor’s future M&A and adopt a point process.

3 Wave-like contagious M&A modeling

3.1 Brief on Hawkes processes

Point process and its intensity function Point processes are widely used to model the occurrences of events. In general, a point process is an event sequence $\{e_1, \dots, e_n\}$ with timestamp $\mathcal{T} = \{t_1, \dots, t_n\}$. Denote $N(t)$ the number of occurred events before t , and $\mathcal{H}_t = \{e_i | t_i < t\}$ as the past events before t . The main concept of point process is the conditional intensity function, or intensity function for short, is given by:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{E}(N(t + \Delta t) - N(t) | \mathcal{H}_t)}{\Delta t} = \frac{\mathbb{E}(dN(t) | \mathcal{H}_t)}{dt}$$

where $\mathbb{E}(dN(t) | \mathcal{H}_t)$ is the expectation of the number of events happened in the interval $[t, t + \Delta t]$ given the historical observations \mathcal{H}_t . The conditional intensity function represents the expected instantaneous rate of events at time t .

Hawkes processes For a self-exciting Hawkes point process, its conditional intensity is written by [Hawkes, 1971]:

$$\lambda(t) = \mu + a \sum_{i: t_i < t} g(t - t_i),$$

where μ is the *spontaneous* intensity and t_i the timestamp of event e_i in the process before time t . $g(t)$ models the exciting effect from the previous events. One general joint likelihood of observing a sequence of events $\mathcal{T} = \{t_1, \dots, t_n\}$ within the time window $[0, T]$, can be given as follows:

$$\mathcal{L}(\mathcal{T}) = \prod_{t_i \in \mathcal{T}} \lambda(t_i) \cdot \exp\left(-\int_0^T \lambda(\tau) d\tau\right),$$

3.2 Problem formulation

We formulate the M&A activities as an inhomogeneous Poisson process with the intensity as the sum of a spontaneous intensity modulated by firms’ profiles and a exciting term related to its past activities and influence from other firms.

Maximum-likelihood estimation In the presence of multiple investors, we want to capture not only the self-excitation of behaviors but also the interaction over investors. The intensity of M&A for an investor d is thus given by:

$$\lambda_d(t) = \mu_d + \sum_{i: t_i < t} a_{dd_i} g_{dd_i}(t - t_i), \quad (1)$$

where the spontaneous term μ_d incorporates the inherent tendency of an investor – active investor making more investments benchmarked by a setting without external excitement. a_{dd_i} measures the influence from dimension d_i to d . Specifically for M&A, d_i to d refer to two investors and a_{dd_i} quantifies the impact from investor d_i ’s M&A event to the other investor d ’s. The impact is controlled by the decay function g_{dd_i} whose input usually is the time interval from previous event timestamp t_i to current time t . An exponential time-decaying function is used in the paper: $g_{ij}(t_i - t_j) = w \cdot e^{-w(t_i - t_j)}$ for its wide popularity and efficacy.

Then for a M&A event sequence $\mathcal{T} = \{(t_i, d_i)\}_{i=1}^n$ of time t_i associated with investor d_i , the log-likelihood is:

$$\mathcal{L}_{cond} = \sum_{d=1}^D \left\{ \sum_{(t_i, d_i) \in \mathcal{T} | d_i = d} \log \lambda_{d_i}(t_i) - \int_0^T \lambda_d(t) dt \right\},$$

By plugging Eq. 1, we obtain the following objective function [Liniger, 2009] where $G_{dd_j}(t) = \int_0^t g_{dd_j}(t) dt$.

$$\mathcal{L}_{cond} = \sum_{i=1}^n \log \left(\mu_{d_i} + \sum_{t_j < t_i} a_{d_i d_j} g_{d_i d_j}(t_i - t_j) \right) - T \sum_{d=1}^D \mu_d - \sum_{d=1}^D \sum_{j=1}^n a_{dd_j} G_{dd_j}(T - t_j) \quad (2)$$

Gaussian prior on spontaneous term So far, at the first glance, the multivariate Hawkes Process seems able to model self and mutual interactions among investors. However, direct application to M&A prediction will incur two major issues. First, the model only takes past events into consideration and neglects the intrinsic characteristic for an individual investor that can affect the M&A action. Second, assigning each company or investor a customized spontaneous term brings the burden of massive parameters to learn and higher risk of overfitting. However, it is also unrealistic to enforce all investors to share one common parameter μ_d . It is appealing to parameterize the spontaneous term μ_d via the profile covariates associated with a company e.g. company size, financial assets etc. Involving such covariates will also increase the model’s interpretability and help identify the influential factors.

Without loss of generality, here we concretely use a Logistic regression function by $\{\beta / (1 + \exp(-\theta^T \mathbf{x}^d))\}_{d=1}^D$ to model the spontaneous intensity term for each investor d where $\mathbf{x}^d = [1, x_1^d, x_2^d, \dots, x_K^d]^T$ concatenate the attributes as summarized in Table 2 and topic features in Table 1. Accordingly, $\theta = [\theta_0, \theta_1, \dots, \theta_K]^T$ are the coefficients, and β is a scaling factor. Rather than μ_d , θ and β are the model parameters needing to be learned from the data.

We adopt a probabilistic view on spontaneous term: a Gaussian prior $\mu_d \sim \mathcal{N}(\beta / (1 + \exp(-\theta^T \mathbf{x}^d)), \frac{1}{2\beta\rho})$ for the regression value is used, which is mathematically equivalent

Algorithm 1 Profile-specific Multi-dimensional Hawkes Process (PMHP) learning for M&A modeling

- 1: Input: M&A event sequence $\{t_i, d_i\}_{i=1}^n$ associated with time t_i and investor i.e. dimension d_i for each event; Profile $\mathbf{x}^d = [x_1, \dots, x_K]^T$ for an investor d ;
 - 2: Initialization for $\{\beta, \rho\}$, $\{\theta_k\}_{k=0}^K$, $\{\mu_d\}_{d=1}^D$, $\{a_{uv}\}_{u,v=1}^D$;
 - 3: **while** Not converge or $l < L$ **do**
 - 4: Update $\{p_{ii}^{(l+1)}\}$ and $\{p_{ij}^{(l+1)}\}$ by Eq.5 and Eq.6;
 - 5: Update $\{\mu_d^{(l+1)}\}_{d=1}^D$, $\{a_{ij}^{(l+1)}\}_{i,j=1}^D$ by Eq.7, Eq.8;
 - 6: Update $\{\beta^{(l+1)}, \rho^{(l+1)}\}$ by Eq.10, Eq.11;
 - 7: Update $\{\theta_k^{(l+1)}\}_{k=0}^K$ by gradient descent in Eq.12;
 - 8: **end while**
-

to measure the deviation by: $(\mu_d - \beta / (1 + \exp(-\theta^T \mathbf{x}^d)))^2$. Its log-likelihood for dimension d is $\mathcal{L}_{pri} = -\frac{\rho}{2} (\mu_d - \beta / (1 + \exp(-\theta^T \mathbf{x}^d)))^2 + \frac{1}{2} \log(\rho)$.

By incorporating both the conditional and prior terms as discussed above, the overall posterior maximum-likelihood estimation problem can be written by $\mathcal{L}(\Theta, \beta, \theta) = \mathcal{L}_{cond} + \mathcal{L}_{pri}$, $\Theta = \{\mu, \mathbf{a}\}$ is introduced for convenience:

$$\begin{aligned} \mathcal{L}_{cond} &= \sum_{i=1}^n \log \left(\mu_{d_i} + \sum_{t_j < t_i} a_{d_i d_j} g_{d_i d_j}(t_i - t_j) \right) \quad (3) \\ &\quad - T \sum_{d=1}^D \mu_d - \sum_{d=1}^D \sum_{j=1}^n a_{dd_j} G_{dd_j}(T - t_j) \\ \mathcal{L}_{pri} &= -\frac{\rho}{2} \sum_{d=1}^D \left\| \mu_d - \frac{\beta}{1 + \exp(-\theta^T \mathbf{x}^d)} \right\|_2^2 + \frac{D}{2} \log(\rho) \end{aligned}$$

Given the learned model parameters, the prediction score s_d for an investor d who might have M&A in a certain future time $[T, T + \Delta T]$ is given by firstly simulating $\lambda_d(t)$ of Hawkes process [Dassios and Zhao, 2013] on time interval $[T, T + \Delta T]$ and then calculating the conditional cumulative distribution by integrating the intensity function.

3.3 Learning Algorithm

To efficiently solve the resulting optimization problem, we design an algorithm which combines techniques of alternating optimization and Majorize-Minimization algorithm [Hunter and Lange, 2004] to maximize $\mathcal{L}(\Theta, \beta, \theta)$. The optimization task is summarized in Algorithm 1. We provide the algorithm step-by-step details as follows.

Finding the lower bound as surrogate

Since we consider an exciting reciprocal affection instead of an inhibiting one, thus the coefficients shall be nonnegative $a_{ij} \geq 0$. This important property leads to a result that $\mathcal{L}(\Theta, \beta, \theta)$ can be surrogated by its lower bound

$\bar{\mathcal{L}}(\Theta|\Theta^{(l)}, \beta, \theta)$ via Jensen's inequality:

$$\begin{aligned} \bar{\mathcal{L}} &= \underbrace{\sum_{i=1}^n \left(p_{ii} \log \frac{\mu_{d_i}}{p_{ii}} + \sum_{j=1}^{i-1} p_{ij} \log \frac{a_{d_i d_j} g_{d_i d_j}(t_i - t_j)}{p_{ij}} \right)}_{\text{part of lower bound by Jensen's inequality}} \quad (4) \\ &\quad - \underbrace{\sum_{d=1}^D T \mu_d - \sum_{d=1}^D \sum_{j=1}^n a_{dd_j} G_{dd_j}(T - t_j)}_{\text{part in original form}} + \mathcal{L}_{pri} \end{aligned}$$

where p_{ii}, p_{ij} are defined as follows in the $l + 1$ -th iteration, which involves variables in Θ in the l -th iteration:

$$p_{ii}^{(l+1)} = \frac{\mu_{d_i}^{(l)}}{\mu_{d_i}^{(l)} + \sum_{j=1}^{i-1} a_{d_i d_j}^{(l)} g_{d_i d_j}(t_i - t_j)} \quad (5)$$

$$p_{ij}^{(l+1)} = \frac{a_{d_i d_j}^{(l)} g_{d_i d_j}(t_i - t_j)}{\mu_{d_i}^{(l)} + \sum_{j=1}^{i-1} a_{d_i d_j}^{(l)} g_{d_i d_j}(t_i - t_j)} \quad (6)$$

p_{ij} can be interpreted as the likelihood that the i -th M&A (d_i, t_i) is affected by the previous j -th one (d_j, t_j). p_{ii} is the probability that i -th event is affected by the spontaneous term.

To verify the feasibility of using the above surrogate function, one can find the equation holds if and only: $\Theta = \Theta^{(l)}$ since p_{ii} and p_{ij} are a function w.r.t. $\Theta^{(l)}$ in the $l + 1$ th iteration. As a result, we have the following relation:

$$\begin{aligned} \mathcal{L}(\Theta, \beta, \theta) &\geq \bar{\mathcal{L}}(\Theta|\Theta^{(l)}, \beta, \theta) \\ \mathcal{L}(\Theta^{(l)}, \beta^{(l)}, \theta^{(l)}) &= \bar{\mathcal{L}}(\Theta^{(l)}|\Theta^{(l)}, \beta^{(l)}, \theta^{(l)}) \end{aligned}$$

Moreover, since the variables β, θ are separable from $\mathcal{L}, \bar{\mathcal{L}}$, we can use the following relation without involving β, θ :

$$\mathcal{L}(\Theta^{(l)}) = \bar{\mathcal{L}}(\Theta^{(l)}|\Theta^{(l)}) \geq \bar{\mathcal{L}}(\Theta^{(l+1)}|\Theta^{(l)}) \geq \mathcal{L}(\Theta^{(l+1)})$$

This implies maximizing $\bar{\mathcal{L}}$ at each iteration ensures that the value of \mathcal{L} increase monotonically.

Solving μ, \mathbf{a} by fixing profile parameters β, θ

The advantage of using the surrogate function is that the parameter μ, a_{ij} can be solved in closed forms, and the nonnegativity constraint of μ, a_{ij} is automatically satisfied.

Zeroing partial derivatives $\frac{\partial \bar{\mathcal{L}}}{\partial \mu_d}, \{\frac{\partial \bar{\mathcal{L}}}{\partial a_{uv}}\}_{u,v=1}^D$ leads to:

$$\mu_d^{(l+1)} = \frac{-h(\mathbf{x}^d) + \sqrt{h(\mathbf{x}^d)^2 + 4\rho^{(l)} \sum_{i=1, d_i=u}^n p_{ii}^{(l+1)}}}{2\rho^{(l)}} \quad (7)$$

$$a_{uv}^{(l+1)} = \frac{\sum_{i=1, d_i=u}^n \sum_{j=1, d_j=v}^{i-1} p_{ij}^{(l+1)}}{\sum_{j=1, d_j=v}^n G_{uv}(T - t_j)} \quad (8)$$

where we define $h(\mathbf{x}^d) = T - \rho^{(l)} \frac{\beta^{(l)}}{1 + \exp(-\theta^{(l)T} \mathbf{x}^d)}$.

$$\mathcal{L}_{pri} = \frac{\rho^{(l)}}{-2} \sum_{d=1}^D \left\| \mu_d^{(l+1)} - \frac{\beta^{(l)}}{1 + \exp(-\theta^{(l)T} \mathbf{x}^d)} \right\|_2^2 + \frac{D}{2} \log \rho^{(l)} \quad (9)$$

Zeroing derivatives $\frac{\partial \mathcal{L}}{\partial \rho}, \frac{\partial \mathcal{L}}{\partial \beta}$ leads to:

$$\beta^{(l+1)} = \frac{\sum_{d=1}^D \mu_d^{(l+1)}}{\sum_{d=1}^D \frac{1}{1 + \exp(-\theta^{(l)T} \mathbf{x}^d)}} \quad (10)$$

$$\rho^{(l+1)} = \frac{1}{D} \sum_{d=1}^D \left(\mu_d^{(l+1)} - \beta / \left(1 + \exp(-\theta^{(l)T} \mathbf{x}^d) \right) \right)^2 \quad (11)$$

Table 1: Top words for each topic learned from news articles by LDA: Top 20 words are listed for each topic.

Topic 1	Software services management information security enterprise business data service product cloud technology financial manage application provider support platform system global
Topic 2	Company inc online top acquired united family credit payment e-commerce canada new site acquisition name service small financial travel international
Topic 3	Digital content entertainment video platform consumer time virtual product social create people around buy world use experience leader strategy 3d
Topic 4	Health care company medical treatment pharmaceutical inc development business range services breast include segment used major consumer delivery production clinical
Topic 5	Million global world leading business portfolio new capital technology leader group billion company equity around investment across people private innovative

Table 2: Investor firm profile covariates from CrunchBase. **VCI** VC, **PE** investments, **IPR** investors per round, **PFB** team members with finance background, **KFB** key person with finance background, **MAY** mean of acquisitions/year, **VAY** variance acquisition # per year, **AMA** amount per acquisition.

Covariate	Mean	Std.	Covariate	Mean	Std.
Investment #	3.8	10.0	Office #	2.31	4.23
Acquisition #	11.83	18.2	Team #	21.7	41.0
Board members #	4.8	5.6	VCI #	0.9	1.8
Employee #	10M	46M	IPR #	1.53	1.08
Amount per round	56M	345M	PFB #	0.06	0.2
Competitor acquired #	0.38	0.88	News #	91.8	506.0
Competitor #	2.7	5.5	KFB #	1.4	1.9
Investor #	1.9	4.6	MAY #	2.04	1.19
Product #	3.56	13.90	VAY	1.68	5.46
Funding round #	1.33	2.28	AMA	616M	1630M

We apply gradient ascent to update $\{\theta_k\}_{k=1}^K$:

$$\frac{\partial \mathcal{L}}{\partial \theta_k} = \rho \sum_{d=1}^D \left(\mu_d - \frac{\beta}{1 + \exp(-\theta^T \mathbf{x}^d)} \right) \frac{\beta \mathbf{x}_k^d \exp(-\theta^T \mathbf{x}^d)}{1 + \exp(-\theta^T \mathbf{x}^d)} \quad (12)$$

4 Experiments on CrunchBase

CrunchBase dataset The dataset is TechCrunch’s open database with information about startups, investors, trends, companies etc. The CrunchBase allows public access to its data via JSON API, by which we’ve collected a local copy of the data as of May 2015. Our used subset consists of 413 companies by two filter criterion: i) the primary role is company; ii) the number of historical M&A is more than 4. We believe this preprocessing can help filter out those dormant players. In this subset, we note several major brands e.g. IBM, Google, Yahoo, Dell etc, associated with profile tags e.g. company size, funds, news articles – see Table 2.

We also consider news reports as they discuss emerging technologies, products, new trends and sometimes acquisition rumors. Previous researches [Xiang *et al.*, 2012] have also shown that text analysis benefits acquisition prediction. We employ Latent Dirichlet Allocation (LDA) to derive the topical features and set the number of topics to be 5, resulting a distribution of an underlying set of topics for each company. Table 1 shows the top 20 words in each topic. The final profile input is a concatenation of profile and topic covariates.

In general, our dataset is comprehensive, free, and up-to-date. In contrast, previous studies e.g. [Pasiouras and Gaganis, 2007] use financial data predominantly while [Grinblatt and Keloharju, 2000] focus on investments in Finland only. **Compared methods** Peer methods are: i) regression models: Linear Regression (LR), Support Vector Machine (SVM) for

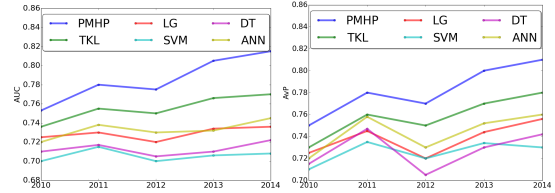


Figure 2: AUC and AvP for prediction year for 2010 – 2014.

Regression, Decision Tree (DT), Artificial Neural Network (ANN); ii) point process: Triggering Kernel Learning (TKL) [Zhou *et al.*, 2013b] that models the spontaneous rate without profiles. Hence their learning algorithm is also different from ours. Note our method is termed Profile-specific Multi-dimensional Hawkes Process (PMHP) in Fig.3 and Fig.4.

Experiments protocol The prediction is performed on a rolling basis: We collect all M&A events and company profiles till year $t - 1$ to predict the M&A intensity over the next year i.e. year t . For point process models, no target supervision is needed and all the data before year t is used to train the models. While regression models by nature need supervision telling the number of M&A events in a forward time window. To predict year t ’s M&A, we use the outcome in year $t - 1$ to set the target variable. Data before year $t - 1$ is used to derive the input features for regression models. Features used for regression models includes all covariates in PMPH plus recent acquisition numbers accounting for recency effect captured by Hawkes model dynamically. As depicted in Fig.2, t is set to the first day of from 2010 to 2014 respectively.

For the point process models i.e. our methods and TKL, the prediction score for an investor is calculated by the simulation method for Hawkes process [Dassios and Zhao, 2013], integrated over the period of the prediction year. While for the regression models, their output score directly indicates the propensity of M&A events in the target year.

Hyper Parameter Setting: PMHP uses one hyper parameters, the exponential decay kernel parameter w in $g(t) = w \exp(-w * t)$. For the exponential decay kernel parameter we tested 0.1, 0.5, 1, 2, 5, 10. After iterations of rigorous experiments, we chose 2. The criteria for choosing the optimal value was to maximize the model’s likelihood \mathcal{L}_{cond} .

In line with [Yan *et al.*, 2012; 2013; 2015; Eugene and Yuan, 2013], we use the area under the Receiver Operating Characteristic curve a.k.a. AUC to assess the predictive models. Average precision(AvP) is also adopted to demonstrate the robustness of the proposed model. To compute AUC and AvP, we create the binary label by setting a company instance as positive if it has at least one M&A in the target year t .

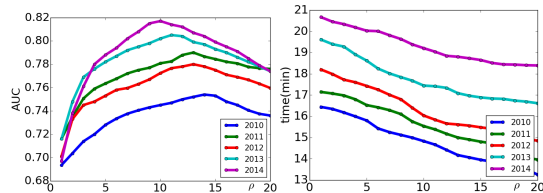


Figure 3: Sensitivity test for ρ by different prediction years.

Prediction performance There are several observations based on Fig.2 and Fig.3: i) PMHP outperforms other methods notably. We conjecture this is because on one hand, our model better captures the wave-like behavior than the regression models limited by aggregated frequency covariates; On the other hand, compared with TKL, our method further explore the profile covariates. This fact is also suggested in recent loosely related work [Guo *et al.*, 2015] in social interaction analysis, where combining content and dynamic behavior information help discover more knowledge than using one of them. ii) As the prediction year moves forward, our method and TKL all show better synergetic effect with the expanded observation time window. This also suggests the advantage of point process models in exploring the dynamic information. iii) Sensitivity test on ρ in Fig.3 indicates PMHP can work in a wide range of this parameter e.g. in the range of [5, 20]. This suggests the practical utility of our method.

Profile importance analysis From Table 3, features e.g. number of past acquisitions, competitors are significant positive indicators for future acquisition events, which implies that companies faced with fierce competition are more possibly involved in acquisition cases. Among The top-ranked negative coefficients are number of investors, key people with financial background and PE/VC investments. One possible explanation is the Post-merger Performance Puzzle [Schipper and Thompson, 1983; Agrawal and Jaffe, 2000], cost of integration may cause a negative long-run stock returns.

Learned mutually-exciting (contagion) matrix $[a_{ij}]$ To illustrate the mutually-exciting phenomenon among companies and industries, in Fig.4 we visualize the mutual-influential matrix $[a_{ij}]$ of seven major industries as defined in Eq.3. Each node represents one company and the thickness of edges between them denote the strength of exciting effect i.e. $|a_{ij}|$. The average value of influence of intra&inter-industries are shown in Table 4. One can find Software industry exhibits relatively stronger influence to others. Note $a_{ij} \neq a_{ji}$.

As shown from the inferred network, intra-industry exciting effects tend to be more visible than inter-industry ones, which may suggest the fierce competition and contagious effect of M&A within industry. Besides, some industries are closely related than others. For instance, software and semiconductors are interweaved and have a strong coupling effect regarding of M&A events while biology technology is not. This suggests the contagion of inter-industry is more selective. These empirical findings also concur with the empirical study observed by [Öberg and Holtström, 2006].

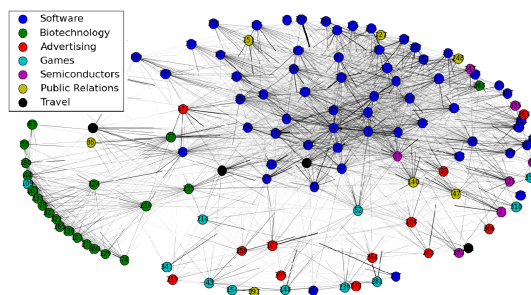


Figure 4: Mutually-exciting phenomenon over firms (nodes) and industries (colors). Edges are the learned mutually-exciting parameter a_{ij} and thicker edges denote higher $|a_{ij}|$.

Table 3: Major covariates and their coefficients learned by our model. Refer to Table 2 for the abbreviations.

	Positive	Coeff.	Negative	Coeff.
	VAY	1.41	Investor #	-0.65
	Team #	1.03	PFB #	-0.59
	AMA	0.76	Investment #	-0.41
	Board member #	0.42	VCI #	-0.37
	Competitor #	0.35	KFB #	-0.33

Table 4: Influential relationships among several industries.

	Software	BioTech	Advertise	Games	Semi conductor	Public Relation	Travel
Software	.0613	.0072	.0137	.0182	.0206	.0122	.0157
BioTech	.0063	.0242	.0039	.0015	.0025	.0023	.0053
Advertising	.0072	.0027	.0168	.0093	.0027	.0048	.0065
Games	.0103	.0015	.0099	.0280	.0047	.0038	.0043
Semiconductor	.0160	.0075	.0034	.0094	.0218	.0055	.0017
Public Relation	.0054	.0047	.0143	.0066	.0074	.0262	.0037
Travel	.0028	.0010	.0062	.0045	.0016	.0127	.0243

5 Conclusion

We propose to use a profile-specific mutually-exciting point process for investors' M&A prediction regardless who is the target company, tailored to the contagious M&A. Results on CrunchBase suggest the suitability of our method to this problem. Future work involves extending the current model and algorithm to other M&A prediction problems e.g. investor-target pair prediction, target identification etc.

References

- [Agrawal and Jaffe, 2000] Anup Agrawal and Jeffrey F Jaffe. The post-merger performance puzzle. *SSRN 199671*, 2000.
- [Ali-Yrkkö *et al.*, 2005] J. Ali-Yrkkö, A. Hyytinen, and M. Pajarinen. Does patenting increase the probability of being acquired? evidence from cross-border and domestic acquisitions. *Applied Financial Economics*, 15(14):1007–1017, 2005.
- [Blundell *et al.*, 2012] C. Blundell, J. Beck, and K. Heller. Modelling reciprocating relationships with hawkes processes. In *NIPS*, 2012.
- [Brealey, 2012] R. Brealey. Principles of corporate finance. *Tata McGraw-Hill Education*, 2012.
- [Dassios and Zhao, 2013] A. Dassios and H. Zhao. Exact simulation of hawkes process with exponentially decaying intensity. *Electronic Communications in Probability*, 2013.

- [Eugene and Yuan, 2012] L. Eugene and S. Yuan. Where's the money? the social behavior of investors in facebook's small world. In *ASONAM*, 2012.
- [Eugene and Yuan, 2013] L. Eugene and S. Yuan. Investors are social animals: Predicting investor behavior using social network features via supervised learning approach. In *MLG*, 2013.
- [Grinblatt and Keloharju, 2000] M. Grinblatt and M. Keloharju. The investment behavior and performance of various investor types: a study of finland's unique data set. *Journal of financial economics*, 55(1):43–67, 2000.
- [Gugler and Konrad, 2002] K. Gugler and K.A. Konrad. Merger target selection and financial structure. *Journal of Modern Accounting and Auditing*, 2002.
- [Guo et al., 2015] F. Guo, C. Blundell, H. Wallach, and K. Heller. The bayesian echo chamber: Modeling social influence via linguistic accommodation. In *AISTATS*, 2015.
- [Hamilton, 1989] J. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- [Harford, 2005] J. Harford. What drives merger waves? *Journal of financial economics*, 2005.
- [Hawkes and Oakes, 1974] A. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 1974.
- [Hawkes, 1971] A. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 1971.
- [Hou et al., 2015] R. Hou, J. Yang, C. Yao, and B. McKelvey. How does competition structure affect industry merger waves? a network analysis perspective. *Physica A: Statistical Mechanics and its Applications*, 429:140–156, 2015.
- [Hunter and Lange, 2004] D. R. Hunter and K. Lange. A tutorial on mm algorithms. *Amer. Statist.*, pages 30–37, 2004.
- [Jovanovic and Rousseau, 2001] B. Jovanovic and P. Rousseau. Mergers and technological change: 1885–1998. *Unpublished working paper, Vanderbilt University*, 2001.
- [Jovanovic and Rousseau, 2002] B. Jovanovic and P. Rousseau. The q-theory of mergers. Technical report, National Bureau of Economic Research, 2002.
- [Liniger, 2009] T. J. Liniger. Multivariate hawkes processes. *PhD thesis, Swiss Federal Institute Of Technology, Zurich*, 2009.
- [Maksimovic et al., 2013] Vojislav Maksimovic, Gordon Phillips, and Liu Yang. Private and public merger waves. *The Journal of Finance*, 68(5):2177–2217, 2013.
- [Mariana, 2012] V. Mariana. An overview on the determinants of mergers and acquisitions waves. *Annals of Faculty of Economics*, 1(2):390–397, 2012.
- [Mitchell and Mulherin, 1996] M. Mitchell and J. Mulherin. The impact of industry shocks on takeover and restructuring activity. *Journal of financial economics*, 41(2):193–229, 1996.
- [Nelson, 1959] R. Nelson. How does competition structure affect industry merger waves? a network analysis perspective. *Princeton University Press, Princeton, NJ*, pages 1895–1956, 1959.
- [Öberg and Holtström, 2006] C. Öberg and J. Holtström. Are mergers and acquisitions contagious? *Journal of Business Research*, 59(12):1267–1275, 2006.
- [Ogata, 1988] Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.*, 83(401):9–27, 1988.
- [Ogata, 1998] Y. Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50:379–402, 1998.
- [Pasiouras and Gaganis, 2007] F. Pasiouras and C. Gaganis. Financial characteristics of banks involved in acquisitions: evidence from asia. *Applied Financial Economics*, 17(4):329–341, 2007.
- [Ragothaman et al., 2003] Srinivasan Ragothaman, Bijayananda Naik, and Kumoli Ramakrishnan. Predicting corporate acquisitions: An application of uncertain reasoning using rule induction. *Information Systems Frontiers*, 5(4):401–412, 2003.
- [Resende, 2008] M. Resende. Mergers and acquisitions waves in the uk: a markov-switching approach. *Applied Financial Economics*, 18(13):1067–1074, 2008.
- [Schipper and Thompson, 1983] K. Schipper and R. Thompson. Evidence on the capitalized value of merger activity for acquiring firms. *Journal of Financial Economics*, 11(1):85–119, 1983.
- [Shleifer and Vishny, 2003] A. Shleifer and R. Vishny. Stock market driven acquisitions. *Journal of financial Economics*, 2003.
- [Slowinski et al., 1997] R. Slowinski, C. Zopounidis, and A. Dimitras. Prediction of company acquisition in greece by means of the rough set approach. *European Journal of Operational Research*, 100(1):1–15, 1997.
- [Song and Chu, 2006] X. Song and Y. Chu. A study on financial strategy for determining the target enterprise of merger and acquisition. *Journal of Modern Accounting and Auditing*, 2006.
- [Very et al., 2012] P. Very, E. Metais, S. Lo, and P. Hourquet. Can we predict m&a activity? *Advances in Mergers and Acquisitions*, 11:1–32, 2012.
- [Weber and Dholakia, 2000] J. Weber and U. Dholakia. Including marketing synergy in acquisition analysis:: A step-wise approach. *Industrial Marketing Management*, 2000.
- [Wei et al., 2009] C. Wei, Y. Jiang, and C. Yang. Patent analysis for supporting merger and acquisition (m&a) prediction: A data mining approach. In *Designing E-business systems. Markets, services, and networks*, pages 187–200. 2009.
- [Xiang et al., 2012] G. Xiang, Z. Zheng, M. Wen, J. Hong, C. Rosé, and C. Liu. A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on techcrunch. In *ICWSM*, 2012.
- [Yan et al., 2012] Junchi Yan, Chunhua Tian, Jin Huang, and Yu Wang. Load forecasting using twin gaussian process model. In *SOLI*, 2012.
- [Yan et al., 2013] J. Yan, Y. Wang, K. Zhou, J. Huang, C. Tian, H. Zha, and W. Dong. Towards effective prioritizing water pipe replacement and rehabilitation. In *IJCAI*, 2013.
- [Yan et al., 2015] J. Yan, C. Zhang, H. Zha, M. Gong, C. Sun, J. Huang, S. Chu, and X. Yang. On machine learning towards predictive sles pipeline analytics. In *AAAI*, 2015.
- [Zhou et al., 2013a] K. Zhou, H. Zha, and L. Song. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *AISTATS*, 2013.
- [Zhou et al., 2013b] K. Zhou, H. Zha, and L. Song. Learning triggering kernels for multi-dimensional hawkes processes. In *ICML*, 2013.