Dynamic Analysis of Investor's Community Sentiment: A Hawkes-Process Framework.

DS3: Data Science Summer School, Ecole Polytechnique, Paris, FR - 28 Aug to 1 Sept 2017

Serge DAROLLES [1,2] Gaëlle LE FOL [1,2] Yoann LE NY[1,3] Guillaume MONARCHA[3]

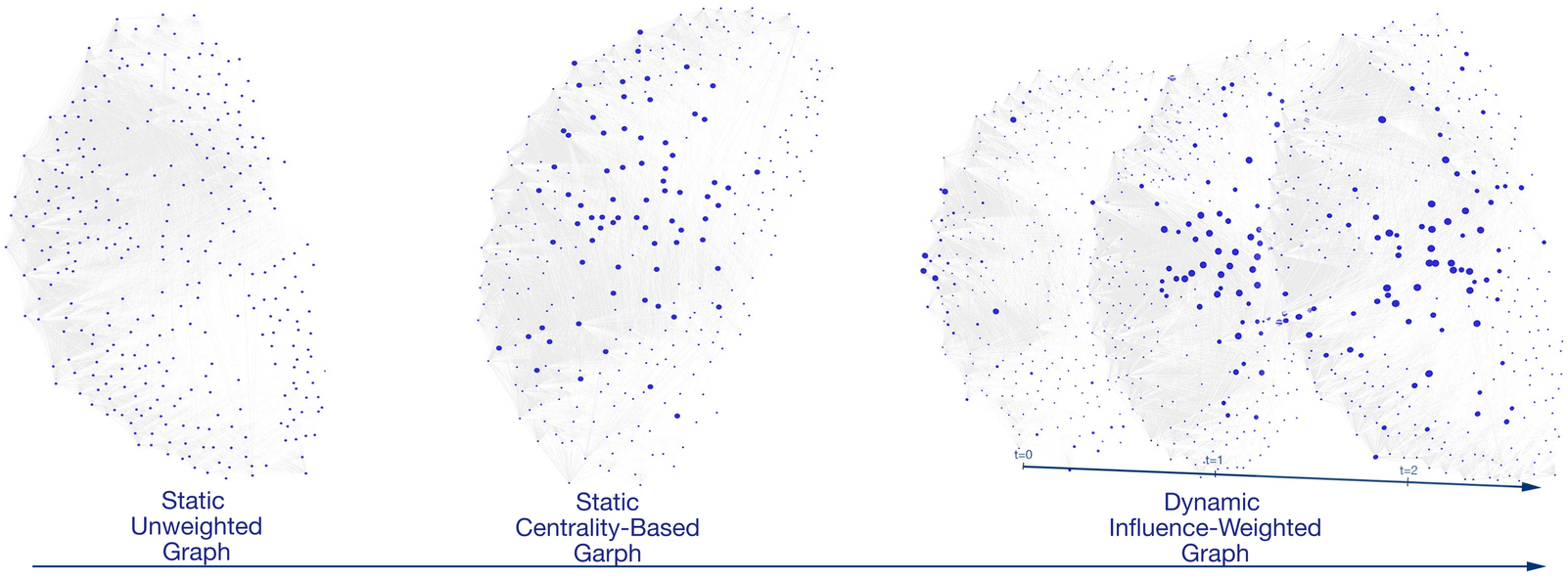
[1]PSL Research-University - Université Paris Dauphine CNRS, Paris, France

[2]INSEE-CREST, Paris, France

[3]Orion Financial Partners, Paris, France

Objectives

This paper aims to exhibit the importance of user's influence on social-network content sentiment extraction. We use a dynamic measure of influence to build dynamic networks of communities. Our final objective is to improve a general measure of the network's sentiment in the form of a time-series index. This measure will be specific to the financial community and tested according to it's significance on financial markets returns predictions.



Introduction

We represent the influence of agents in the financial community as a self-exciting point-process. Contrary to other studies that consider the financial community on social media as a static entity, our dynamic approach reduces the size of the community at each time frame and the noise associated to a too large size. We then extract the sentiment from this dynamic community.

Data

Table: Database descriptive statistics

	31-12-2009 to 31-12-2016
Number of Users	192,486
Number of Twits	105,430,254
Average Daily Volume	48,141

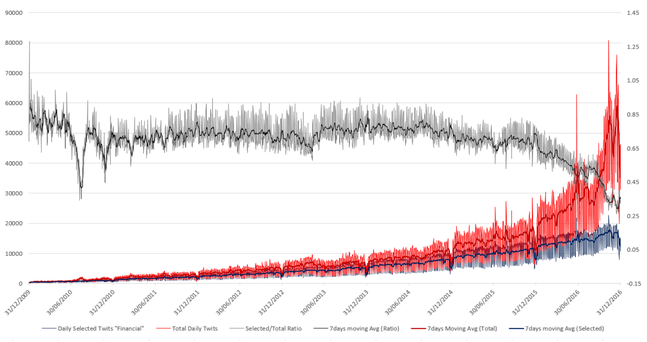
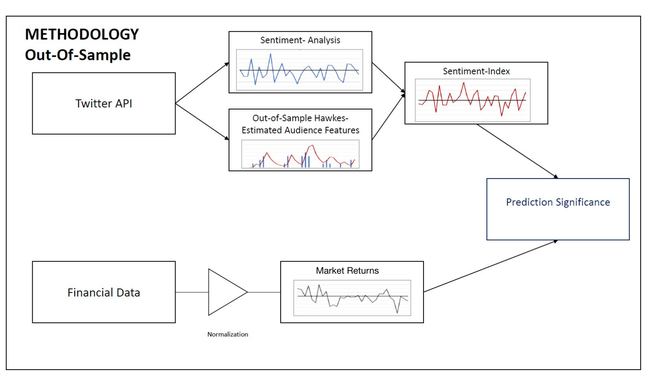


Figure: Content volume from dec-2009 to dec-2016

Methodology

The general process of this work involves first a processing phase including text-analysis and Influence modeling then a testing phase to prove the indicator enhanced predictive power.



Text Analysis

Models:

- Target Extraction: POS tagger
- Target Classification: Conv-Net Based on Word2Vec features
- Sentiment Analysis: Naive-Bayesian classifier trained on a large dataset of 1.6 Million labeled tweets.

Influence Model

With use parametrized Hawkes self-exciting point-process defined with the counting process $(N(t) : t > 0)$ Its intensity following:

$$\lambda(t) = \mu + \int_0^t \alpha(t-u)dN_u$$

Which, using $\alpha(t) = \alpha e^{-\beta t}$, leads to:

$$\lambda_t = \mu + \int_{-\inf}^{t} \alpha e^{-\beta(t-s)} dN_s = \mu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}$$

We estimate parameters by maximizing the Log-likelihood function.

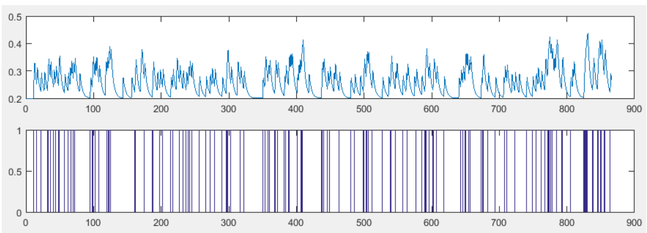


Figure: Hawkes-Process modeling of Audience-Feature *'IN-OUT Ratio'* for user *@Carl_C_Icahn* from 31-dec-2009 to 31-dec-2016 .

Sentiment Index

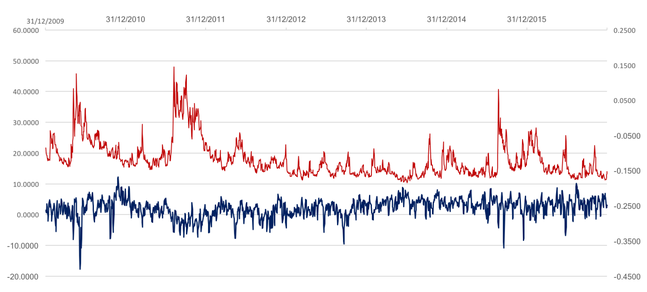


Figure: Sentiment-Index (SentIn) and VIX Index from 31-dec-2009 to 31-dec-2016 .

Results

Linear Estimation

$$R_{t+k} = \alpha + \beta Scomparable_t + \epsilon_t$$

Sentiment Measure	VIX	InSent	Δ VIX	Δ InSent
S&P 500 Index (t+1)				
β	0.0012	0.0026	0.0005	0.0004
(p-value)	**0.00**	0.66	**0.04**	**0.05**
R-squared	0.00	0.00	0.13	0.06

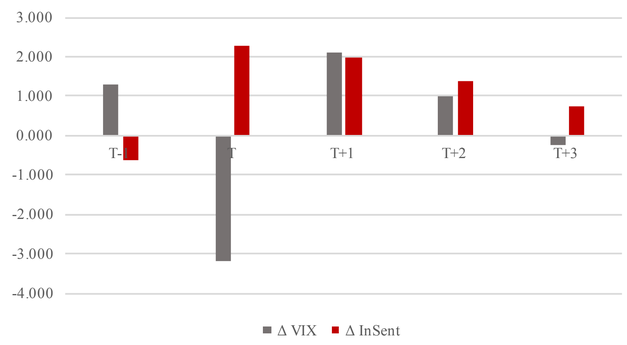


Figure: Lead/Lag Structure .

Two-Factors Model

$$R_{t+k} = \alpha + \beta_1 InSent_t + \beta_2 VIX_t + \epsilon_t,$$

SP500 Idx (t+1)		t-Stat	p-Value
intercept	0.0000	0.00	1.00
ΔVIX	0.0005	2.29	**0.02**
ΔInSent	0.0006	2.76	**0.01**

VAR

$$Y_{1,t} = c + A_{1,1}Y_{1,t-1} + A_{1,2}Y_{2,t-1} + \epsilon_{1,t}$$

$$Y_{2,t} = c + A_{2,1}Y_{1,t-1} + A_{2,2}Y_{2,t-1} + \epsilon_{2,t}$$

$Y_t = \Delta VIX_t$	$\Delta InSent_{t-1}$	ΔVIX_{t-1}	c
Panel A: ΔVIX_t			
(t-stat)	-2.111	-4.357	-0.126
(p-value)	**0.035**	**0.000**	0.900
R-squared	0.09		

$Y_t = \Delta InSent_t$	$\Delta InSent_{t-1}$	ΔVIX_{t-1}	c
Panel B: $\Delta InSent_t$			
(t-stat)	-16.949	-2.355	-0.160
(p-value)	**0.000**	**0.019**	0.873
R-squared	0.21		

Conclusion

- We Use an Hawkes-Process Modeling to provide a dynamic measure of social-network users' influence
- We showed that this Influence metric enhances the significance of the Sentiment signal to predict stock-markets returns
- Provide proofs that this measures provides a supplementary information to predicting market returns than the usually considered *Fear* indices.