

Social Sensing

Wolfgang K. Härdle

Elisabeth Bommers

Ladislaus von Bortkiewicz Chair of Statistics

Humboldt-Universität zu Berlin

lvb.wiwi.hu-berlin.de



σ_t



Internet of Things (IoT)

- ◉ Network between everyday objects
- ◉ Quantified self
- ◉ Apple Watch
 - GPS
 - Daily steps
 - Heart rate
 - Sleep

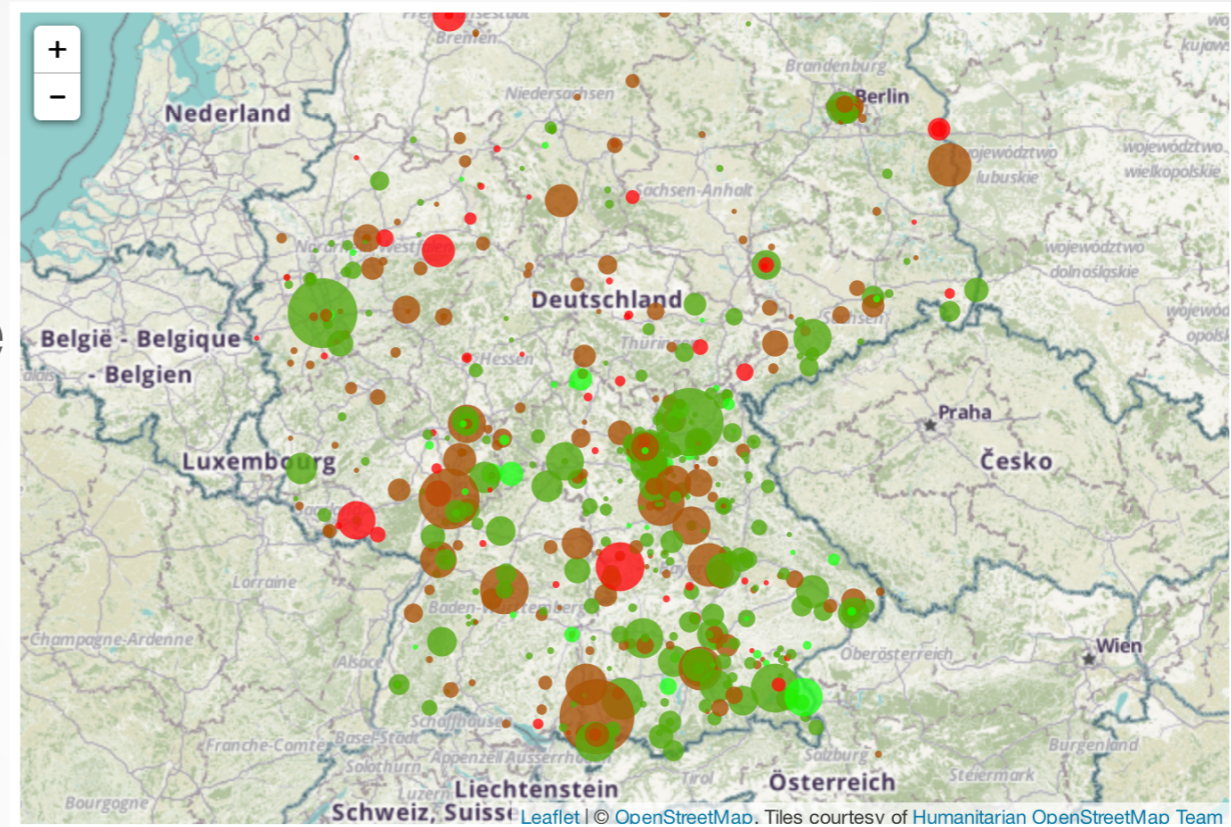


- ◉ How to measure a human's opinion?



German Brewery Ratings

- Numeric score
- BeerAdvocate.com
 - Scrape user ratings
 - Longitude and Latitude
 - Plot interactive map
 - Select best local beer



 TXTBreweries

- Deeper analysis: opportunities and risks of export in other country
- Sentiment often only **implicit** (text) and not explicit (score)



Social Sensing

- Use sentiment numerisization as proxy for sensor

- Sentiment lexica

- Opinion Lexicon* (BL)

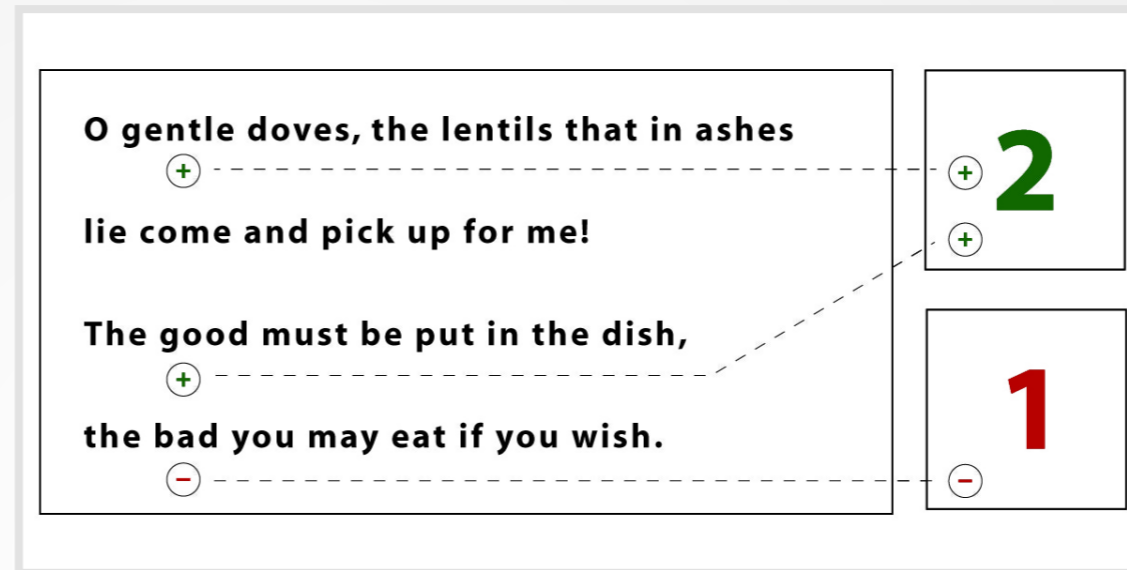
Hu and Liu (2004)

- Financial Sentiment Dictionary* (LM)

Loughran and McDonald (2011)

- Multi-Perspective Question Answering Subjectivity Lexicon* (MPQA)

Wilson et al. (2005)



Workflow



Figure 1

Workflow of Data Gathering, NLP, Sentiment Classification and Modeling



Workflow

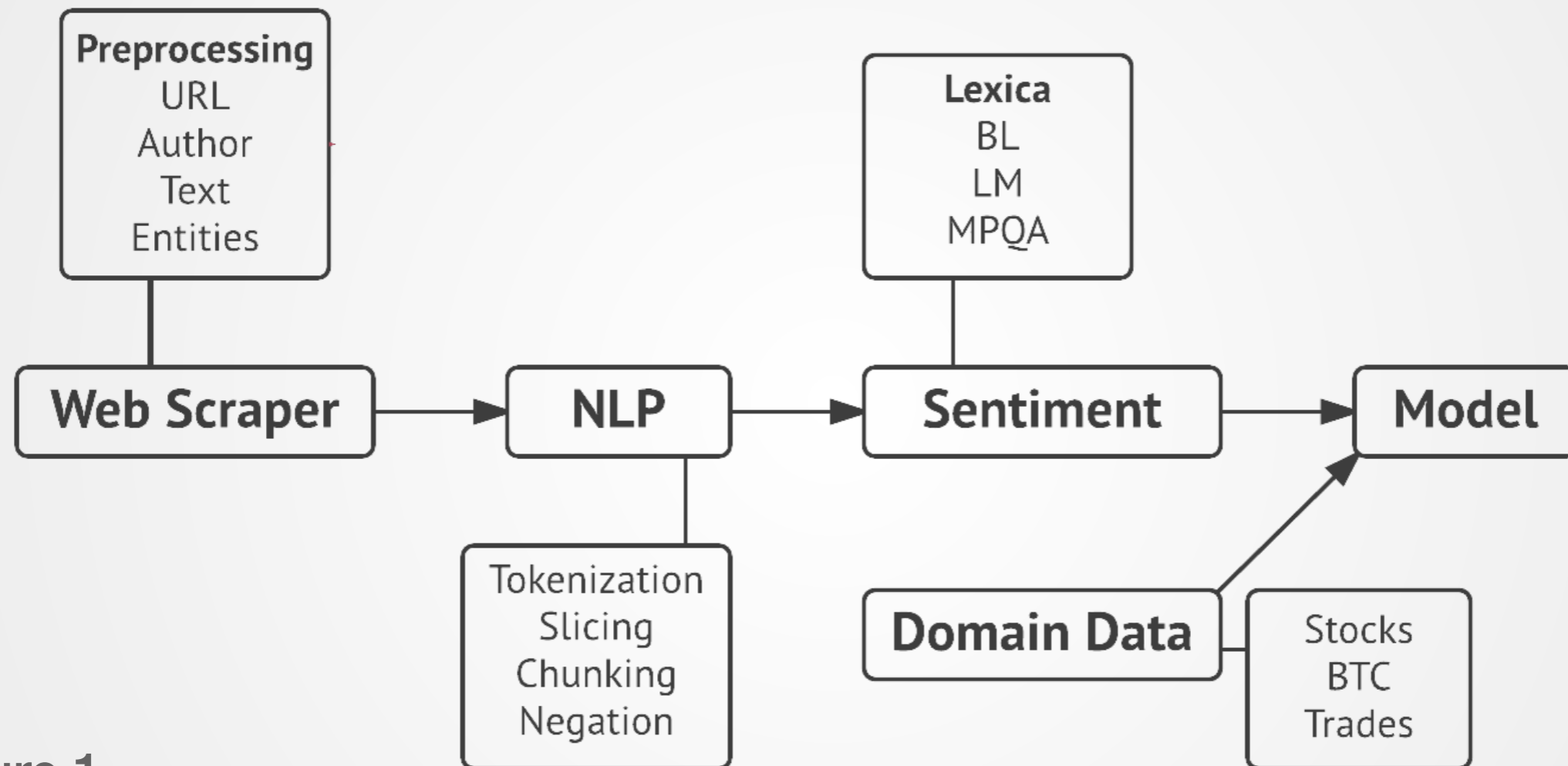


Figure 1

Workflow of Data Gathering, NLP, Sentiment Classification and Modeling



Stocks and Bitcoins

- *Distillation of news flow into analysis of stock reactions*

Zhang, J, Chen, C, Härdle, W K and Bommers, E (2016)

- *Bipeds Bitcoins Blockchain*

Härdle, W K, Teo E G S and Bommers, E (2015)



Conclusion

- Public web-based texts offer new opportunities for research
- Human opinion can be measured
- Insights also interesting for industry
- Domain specific data gathering, processing and analysis
- Data itself is free (moneywise) but time-consuming processing



Social Sensing

Wolfgang K. Härdle

Elisabeth Bommers

Ladislaus von Bortkiewicz Chair of Statistics

Humboldt-Universität zu Berlin

lvb.wiwi.hu-berlin.de



σ_t



Project Idea

- Influence of sentiment in news on stock reaction indicators:
 - Garman-Klaas range-based log volatility ($\log \sigma$)
 - Detrended log trading volume (V)
 - Return (R)
- Nasdaq news articles
 - October 2009 – October 2014
 - 116,691 articles in total
 - 43,459 articles about 100 selected S&P 500 stocks
 - Data available at [RDC](#)



Sentiment Variables

- ◉ $I_{i,t}$ - article indicator
 - ◉ $Pos_{i,t}$ - average proportion of positive words
 - ◉ $Neg_{i,t}$ - average proportion of negative words
- for firm i on day t



Panel Regression

$$\left. \begin{array}{l} \log \sigma_{i,t+1} \\ V_{i,t+1} \\ R_{i,t+1} \end{array} \right\} = \alpha + \beta_1 I_{i,t} + \beta_2 Pos_{i,t} + \beta_3 Neg_{i,t} + \beta_4^\top X_{i,t} + \gamma_i + \varepsilon_{i,t}$$

for stock i on day t with separate estimation of each regression equation

- $X_{i,t}$ - control variables
- γ_i - company specific fixed effect satisfying $\sum_i \gamma_i = 0$



Control Variables

- ◉ $R_{M,t}$ - S&P 500 market returns
- ◉ VIX_t - CBOE VIX
- ◉ $\log \sigma_{i,t}$ - Range-based log volatility
- ◉ $V_{i,t}$ - Detrended log trading volume
- ◉ $R_{i,t}$ - Return



Regression results

Variable	BL	LM	MPQA	PCA
Panel A: Future Log Volatility $\log \sigma_{i,t+1}$				
$I_{i,t}$	-0.005	-0.019***	-0.004	-0.014
$Pos_{i,t}$	-0.396*	0.156	-0.517**	-0.210
$Neg_{i,t}$	0.905***	0.942***	1.464***	1.041***
Panel B: Future Detrended Log Trading Volume $V_{i,t+1}$				
$I_{i,t}$	0.040***	0.027***	0.046***	0.035***
$Pos_{i,t}$	-0.496***	0.051	-0.483**	-0.274*
$Neg_{i,t}$	0.726***	0.563**	0.548*	0.590**
Panel C: Future Returns $R_{i,t+1}$				
$I_{i,t}$	0.000	0.000	0.000	-0.000
$Pos_{i,t}$	0.019***	0.030***	0.014*	0.018***
$Neg_{i,t}$	-0.004	-0.000	-0.009	-0.003

*** p value < 0.01 , ** $0.01 \leq p$ value < 0.05 , * $0.05 \leq p$ value < 0.1



Volatility Simulation

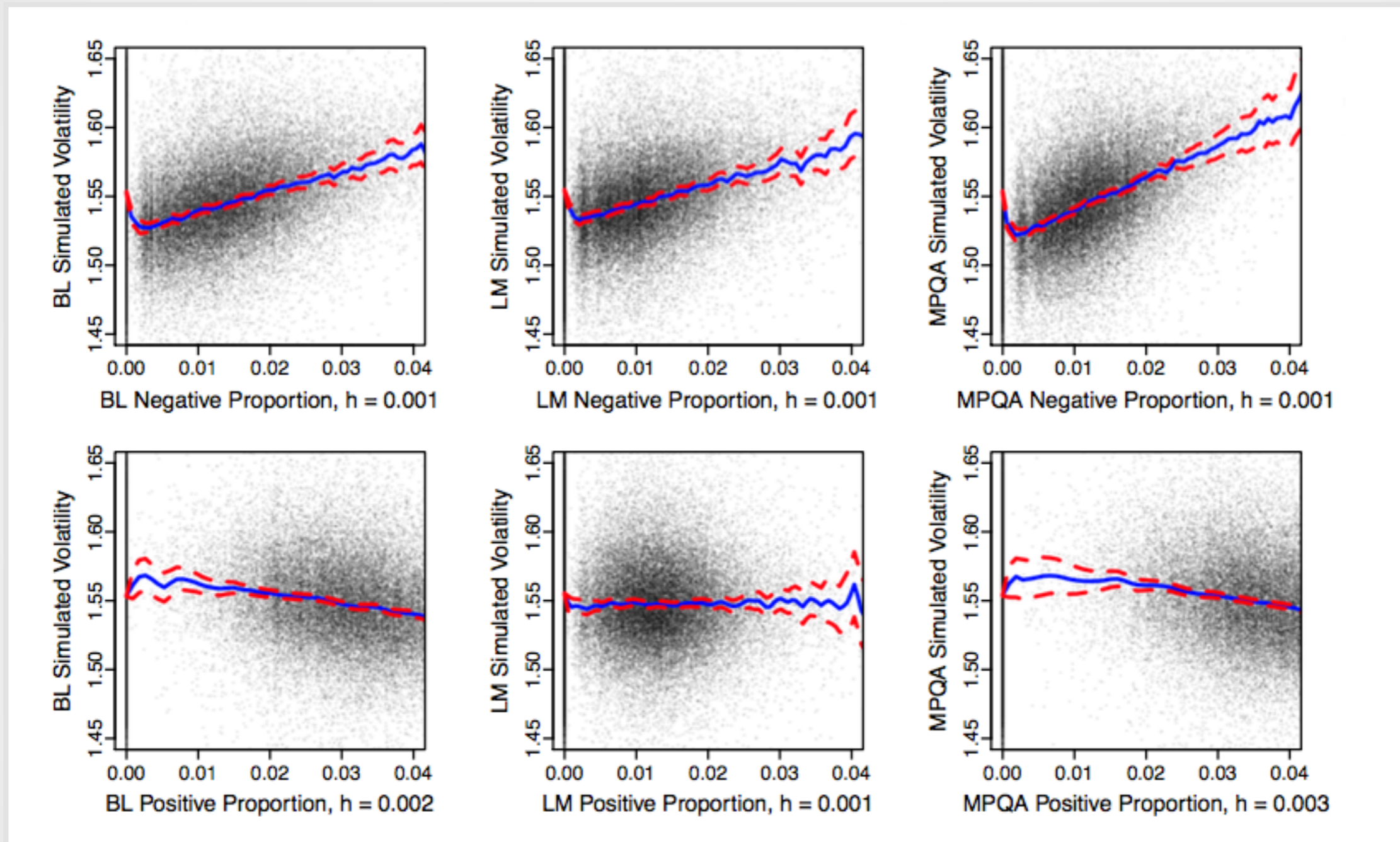


Figure 2

Volatility Simulation for entire panel: mean curve and 95% uniform confidence bands



Conclusion

- Sentiment measures: incremental information about future stock reactions
- Asymmetric impact of positive and negative sentiment
- Degree of incremental information and asymmetry is sector and attention specific
- Choice of lexicon matters

[Back to overview](#)



Bitcoin - Cryptocurrency

- System: decentralized, virtual, low transaction costs



- NYSE, Andreessen Horowitz, DFJ: Coinbase funding (75 M\$)
- Citigroup: own coin development
- Nasdaq: company-wide utilization of **blockchain technology**
 - Main technological innovation: trustless trust
 - Public ledger of all historic transactions

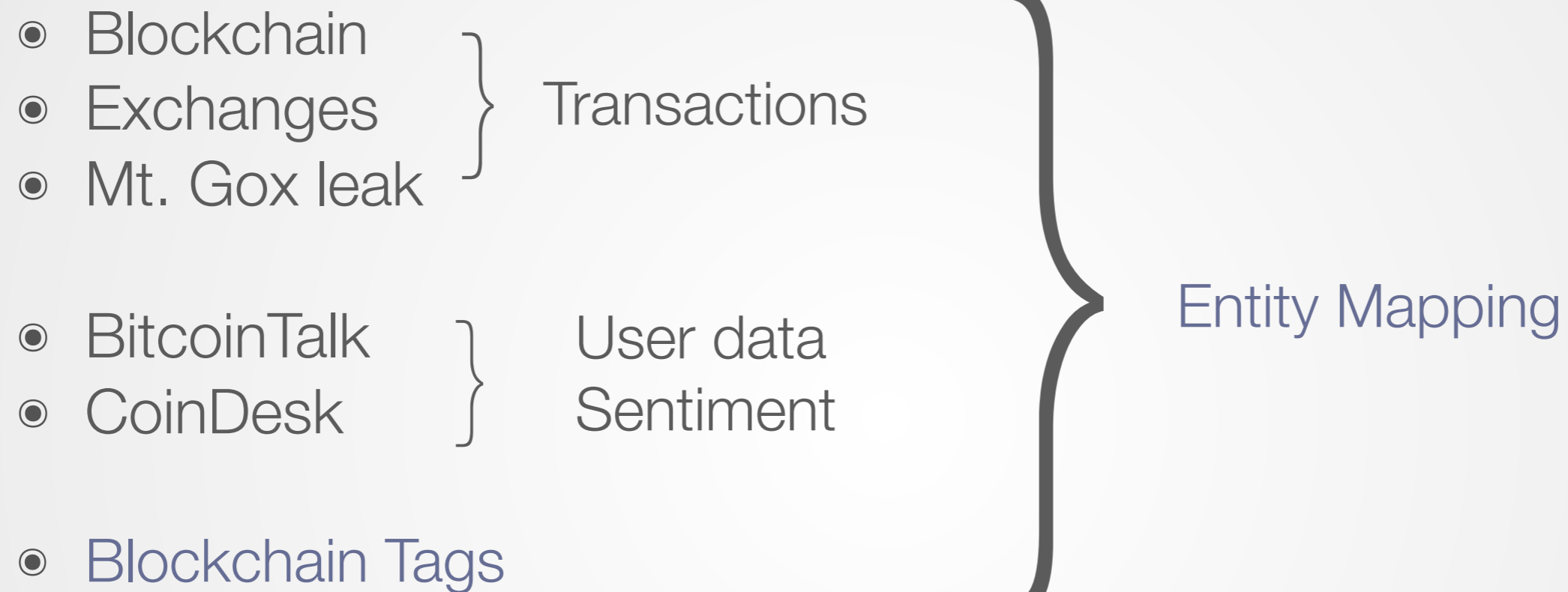


Project Idea

- ◎ Large scale blockchain **graph analysis** (~ 80 GB)
 - 2013 Bitcoin **bubble**
 - **Mt. Gox**: internal fraud or external theft?
 - Development of **risk measure**
 - Possible tools: **spectral clustering**, Majer et al. (2015)
- ◎ Extensive **preprocessing**
 - Obtain, clean, combine data
 - Variable generation (sentiment, attention index)
 - **Trivial?** Not really



Data



Back to overview



References

Hu, M and Liu, B

Mining and Summarizing Customer Reviews
10th ACM SIGKDD, 2004

Loughran, T and McDonald, B

When is a liability not a liability?
J. Financ., 2011

Majer, P, Mohr, P N, Heekeren, H R and Härdle, W K

Portfolio Decisions and Brain Reactions via the CEAD method
Psychometrica, 2015

Wilson, T and Wiebe, J and Hoffmann, P

Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis
HLT-EMNLP, 2005

Zhang, J, Chen, C, Härdle, W K and Bommers, E

Distillation of news flow into analysis of stock reactions
JBES, 2016 (Forthcoming)

All icons by icons8.com

