

## Report on the outcomes of a Short-Term Scientific Mission<sup>1</sup>

**Action number:** CA19130

**Grantee name:** Luciana Dalla Valle

### **Details of the STSM**

Title: Data Integration and Graphical Models for Cryptocurrencies

Start and end date: 11/07/2022 to 18/07/2022

### **Description of the work carried out during the STSM**

Description of the activities carried out during the STSM. Any deviations from the initial working plan shall also be described in this section.

*(max. 500 words)*

The STSM involved a one-week visit of Dr Luciana Dalla Valle, from the University of Plymouth (UK) at the University of Pavia (Italy), to work with Dr Claudia Tarantola.

During the STSM, Dr Dalla Valle and Dr Tarantola discussed the purpose of the research, the methodology to be used, the steps of the analysis to be carried out and the implementation of the computational aspects with the R software.

Dr Dalla Valle and Dr Tarantola opted to focus mainly on the Bitcoin cryptocurrency and decided to explore how the use of social media information may improve market price predictions. The idea is to combine different types of data on Bitcoins using dependence models, such as copulas and vines, and graphical models for time series.

The collaborators performed an extensive literature review, focusing on the analysis of cryptocurrencies, the analysis of textual information, social media sentiment analysis, social network analysis, time series analysis of continuous data, copula dependence models, pair copula constructions, vines and graphical models for time series data.

Dr Dalla Valle and Dr Tarantola collected all the data required for the implementation of the project, choosing as time horizon the period between 06/02/2021 and 15/06/2021.

---

<sup>1</sup> This report is submitted by the grantee to the Action MC for approval and for claiming payment of the awarded grant. The Grant Awarding Coordinator coordinates the evaluation of this report on behalf of the Action MC and instructs the GH for payment of the Grant.

In particular, the data downloaded for the analyses are the following:

- Bitcoin daily prices were downloaded from the website: <https://coinmetrics.io/community-network-data/>; the data description is available at: <https://docs.coinmetrics.io/info/metrics>; the variable to be used for the analysis is called “PriceUSD”, which is defined as “The fixed closing price of the asset as of 00:00 UTC the following day (i.e., midnight UTC of the current day) denominated in USD. This price is generated by Coin Metrics fixing/reference rate service”;
- Google trends containing keywords such as “Bitcoin” and “btp” were downloaded for the selected time horizon using the “gtrendsR” package, available at <https://cran.r-project.org/web/packages/gtrendsR/index.html>;
- Tweets containing hashtags such as “Bitcoin” and “btp” were downloaded from the Kaggle webpage <https://www.kaggle.com/kaushiksuresh147/bitcoin-tweets>.

The downloaded data were cleansed and prepared for the subsequent analyses. In particular, sentiment scores of tweets containing hashtags such as “Bitcoin” and “btp” were calculated using the Bing and Afinn lexicons, using the “tidytext” R package, as illustrated at the link: <https://www.tidytextmining.com/sentiment.html>.

In addition, Dr Dalla Valle and Dr Tarantola started to implement the word network analysis, to analyse the word included in the downloaded tweets using social networks with the aim of identifying associations between words. In particular, the collaborators explored the application to the Bitcoin data of the Textnets R package for automated text analysis using network techniques.

Finally, Dr Dalla Valle and Dr Tarantola developed the R code which implements the first steps of the analysis and drafted out the first version of the research paper describing the methodology.

### **Description of the STSM main achievements and planned follow-up activities**

Description and assessment of whether the STSM achieved its planned goals and expected outcomes, including specific contribution to Action objective and deliverables, or publications resulting from the STSM. Agreed plans for future follow-up collaborations shall also be described in this section.

*(max. 500 words)*

Dr Dalla Valle and Dr Tarantola obtained some very interesting results from the exploratory analysis of the Twitter textual data.

In particular, the collaborators draw maps of the geographical locations of the users who produced Tweets containing the keywords “Bitcoin” and “btp”.

Also, they created lineplots of the number of tweets created during the selected time horizon, comparing them to those produced from the Bitcoin daily prices and Google trends. The plots shows an increasing trends for all time series, suggesting an association between the different types of data.

The sentiment analysis allowed the collaborators to explore the most frequent words used in the collected tweets and to analyse the overall sentiment score in the whole dataset.

In addition the evolution of the sentiment scores was monitored over time, revealing that the trend was similar to the one already suggested by the other Bitcoin time series.

A draft working paper has been written based on the results achieved.

Dr Dalla Valle and Dr Tarantola are planning to present the work carried out at the Round Table “COST Fintech and AI Diversity” of the Women in Fintech II Conference at the University of Tirana on 21-22 September 2022.

The work carried out during the STSM will form the basis of a new research collaboration between Plymouth and Pavia and will be followed by regular online meetings to finalise the implementation of the methodology, complete the paper and submit it to a highly rated international journal.

In particular, the next steps of the analysis, that will be implemented in the next few months, are the following:

- 1) Perform data integration using vine copula specifications.  
First, marginal models will be implemented utilizing time series models for continuous data; then residuals will be extracted and transformed into pseudo-observations to enable copula modelling. Finally, a vine copula model will be estimated, by using the “rvinecopulib” R package, available at <https://cran.r-project.org/web/packages/rvinecopulib/index.html>.
- 2) Perform data integration using graphical models for time series.  
This step will be implemented using the “mgm” R package for estimating time-varying mixed graphical models in high-dimensional data.

Comparison between vine copula and graphical models to predict Bitcoin data.