



EUROPEAN CENTRAL BANK

EUROSYSTEM

Urszula Kochanska

Economist-Statistician

European Central Bank

Directorate General Statistics

Monetary & Economic Statistics (MOE)

Financial Statistics Section (FINS)

Dataset and indicators to monitor the crypto- assets phenomenon

***R. Boumghar, U. Kochanska,
A. Tracz, A. Vouldis, A. Zito***

17 August 2019

The views expressed in this presentation and in the related paper are those of the authors and do not necessarily reflect the views of the European Central Bank.

Any errors and omissions are the sole responsibility of the authors.

Are Crypto Assets Rising ?

1. **Monitoring framework** for crypto-assets based on public data
2. **Indicators** to monitor crypto-asset phenomenon and identified data gaps
3. Overview of **data quality** issues
4. **Cross-mapping and linking crypto-asset data** from various sources
5. Calculating of synthetic exchange rate to express pricing and trading information in **harmonised units** of fiat currencies or crypto-assets
6. Enhancing the quality of pricing and trading information by **applying machine learning techniques** and using alternative data sources
7. Conclusions and way forward

Developments of crypto-assets represents a major **financial innovation with potentially far reaching implications for financial players**

The ECB has been analysing the crypto-asset phenomenon with the view to identifying and monitoring potential implications and risks in the areas of ECB interest.

see e.g.



Part of the analysis consisted of the investigation on the **characterisation of crypto-assets:**

“a new type of assets recorded in digital form and enabled by the use of cryptography that is not and does not represent a financial claim on, or a liability of, any identifiable entity”



Occasional Paper Series

ECB Crypto-Assets Task Force

Crypto-Assets:
Implications for financial stability, monetary policy, and payments and market infrastructures

[Link](#)



Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.



Economic Bulletin

Understanding the crypto-asset phenomenon, its risks and measurement issues (August 2019)

[Link](#)





identification of a range of monitoring **needs**
and review of **the available data**

sources: commercial and non-commercial data providers
granular information allowing the calculation of customised indicators

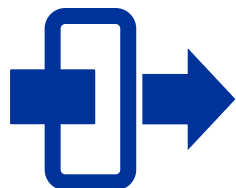
automated procedures to collect information using **API** and **big data** technologies

consistent methodologies, definitions
data quality enhancements

Crypto-asset indicators tailored to the ECB monitoring exercise have been grouped in four categories covering largely off-chain transactions:



i) markets



ii) gatekeepers



iii) linkages



iv) Initial coin offerings (ICOs)

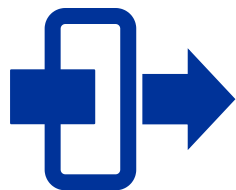
- crypto-asset trading and pricing information gathered from the crypto-assets trading platforms
- market capitalisation of crypto-assets
- trading vis-à-vis fiat currencies
- trading and pricing information of derivatives and ETP with exposures to crypto-assets on institutionalised exchanges

- trading platforms
- wallets supporting crypto-assets
- ATMs supporting crypto-assets
- cards supporting crypto-assets

Crypto-asset indicators tailored to the ECB monitoring exercise have been grouped in four categories covering largely off-chain transactions:



i) markets



ii) gatekeepers



iii) linkages



iv) Initial coin offerings (ICOs)

- linkages of the crypto-asset markets with the financial systems and the real sector of the economy
- holdings of crypto-assets derivatives (bitcoin futures)
- holdings of ETPs offering exposures to crypto-assets

- amounts and features of funds raised e.g. legal form, the underlying blockchain and the country of incorporation.

Blockchain



- transactions
- fees and other



Data gaps

Important data gaps have been identified, particularly relating to certain **interlinkages** and to **payment transactions**, including the use of layered protocols.

Examples:

- amounts of banks and financial corporations' direct exposures or indirect exposures (lending for purposes of investing in crypto-assets)
- transactions with cards supporting crypto-assets
- sales of merchants accepting crypto-assets
- withdrawal transactions from crypto-assets ATMs
- transactions using layered protocols

3. Overview of data quality issues



Data quality of crypto-asset statistics

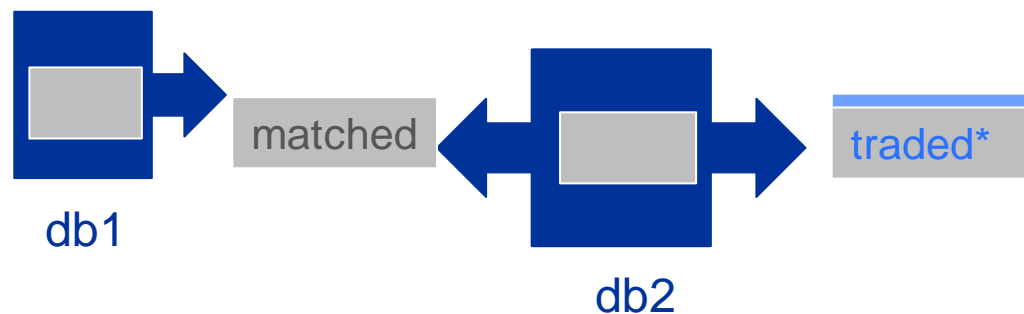
Off-chain transactions

- often publicly available
- analysis can be complex e.g.
 - identifying the value of transactions
 - identifying if different wallets belong to the same entity
- likely more challenging in the future
- layered protocols

On-chain transactions

- wash trading
- fraudulent transactions
- technical issues related to cyberattacks, fraud and hacking
- other

4. Cross-mapping and linking crypto-asset from various sources



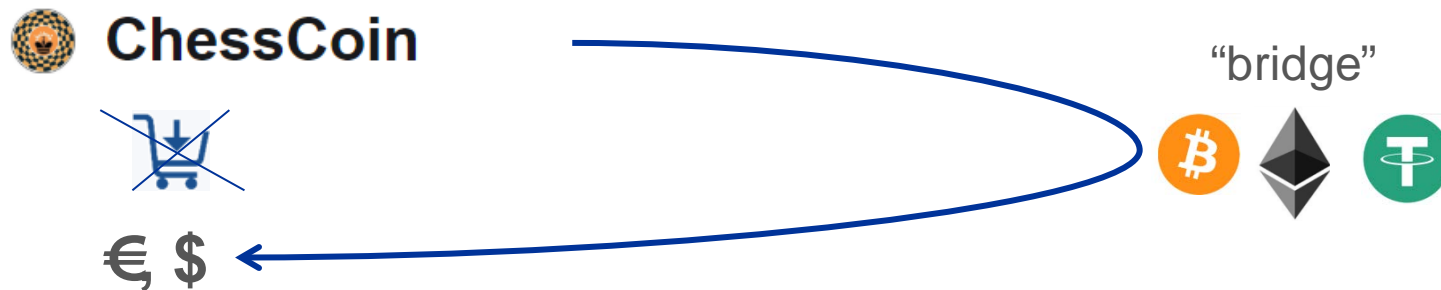
Integration of data from various sources required matching **reference data for crypto-assets**. Rudimental matching (join between symbols) unable to deliver due to differences in symbol used to identify crypto-assets including duplicates

Solution implemented:

- using **string distances measure q-grams** based on the unique identifier created by concatenating symbol and name
 - 70% of db 1
 - 99% of persistently traded crypto-assets (every day since the beginning of 2019)
 - price information was used to validate results

	distance	
CMTcybermiles	0.0000000	CMTcybermiles
CMTcybermiles	0.4395515	CYMTcybermusic
CMTCometcoin	0.2928932	CMTcomet

5. Pricing and trading information in harmonised units



To compare crypto-assets it is required that prices, trading volumes and market capitalisation are expressed in comparable units of fiat currency or crypto-assets e.g. USD, EUR or BTC.

Three **selection criteria** for the bucket of “bridge” crypto-assets

cover 100% of all crypto-assets not trading to fiat currencies

number of “bridge” crypto-assets is the least possible

all “bridge” crypto-assets need to trade to major fiat currencies

Other options under investigation:

- *non-static hierarchy of “bridge” crypto-assets*
- *weighted average price of all vis-à-vis crypto-assets*
- *other*

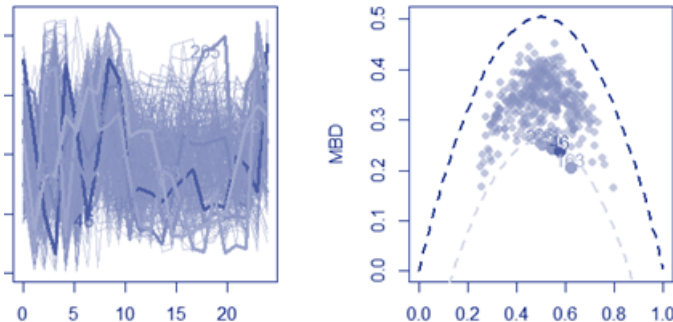
6. Enhancing the quality of pricing and trading information

Objective:

- outlier identification and filtering out
- applying state of the art algorithms of functional data analysis and machine learning using also “alternative” data sources

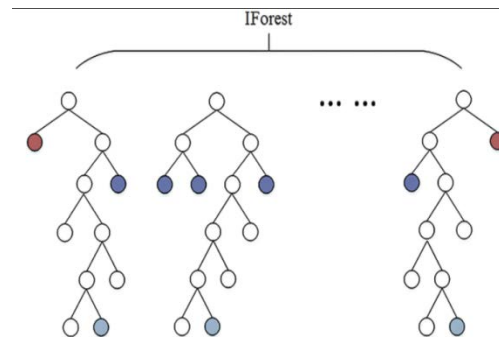
Three approaches

Functional Data Analysis – the Outliergram



Arribas-Gil A., Romo J. (2014)
Shape outlier detection and visualization for functional data: the outliergram
Tarabelloni, N., et al. *roahd: Robust Analysis of High Dimensional Data*. R package version 1 (2017).

Isolation Forest



Liu F.T., Ting K. M., Zhou Z. H. (2008)
Isolation forest
Liu F.T., Ting K. M., Zhou Z. H. (2014)
Isolation-based Anomaly Detection

Dependency Network Analysis

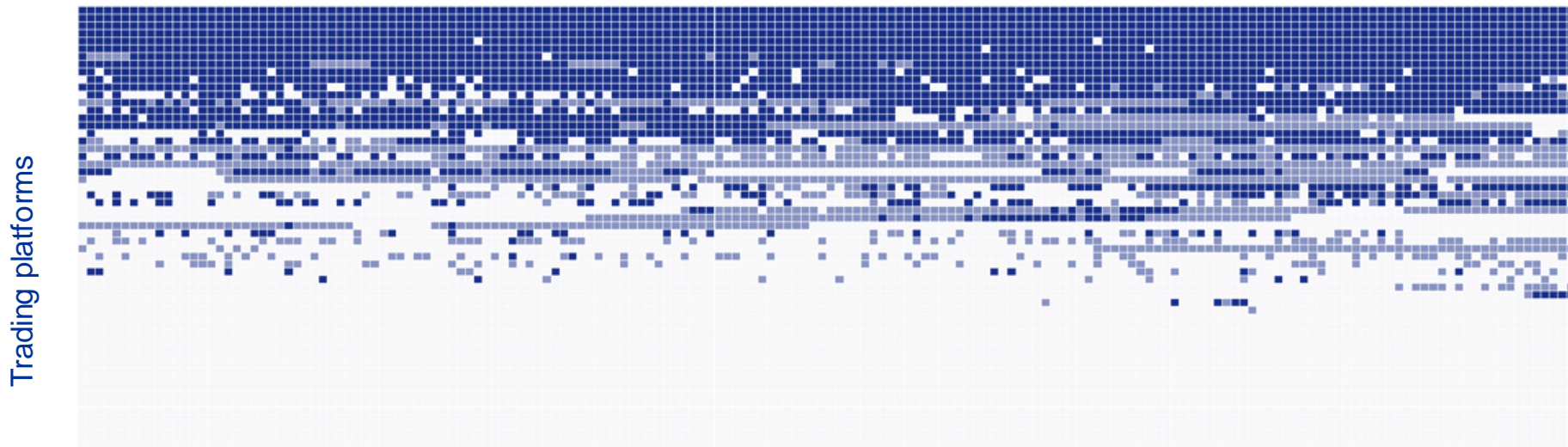
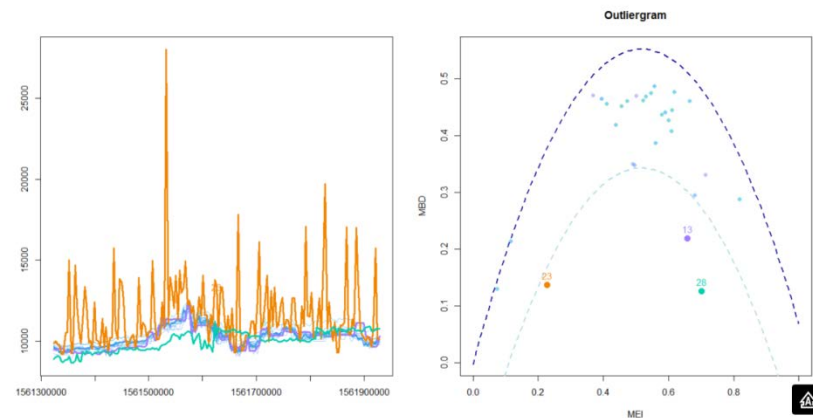


Boumghar R. et al. (2018)
Enhanced Awareness in Space Operations Using Web-Based Interactive Multipurpose Dynamic Network Analysis.
Gu C., Gowda M. G., Boumghar R., Jayanna K. P., et al. (2019)
The Added Value of Advanced Feature Engineering and Selection for Machine Learning Models in Spacecraft Behavior Prediction

6. Enhancing data quality – tool for functional data – the outliergram

Functional Data Analysis – the Outliergram

- method to detect shape outliers in samples of curves
- addressing magnitude outliers and shape outliers



Daily since January 2019

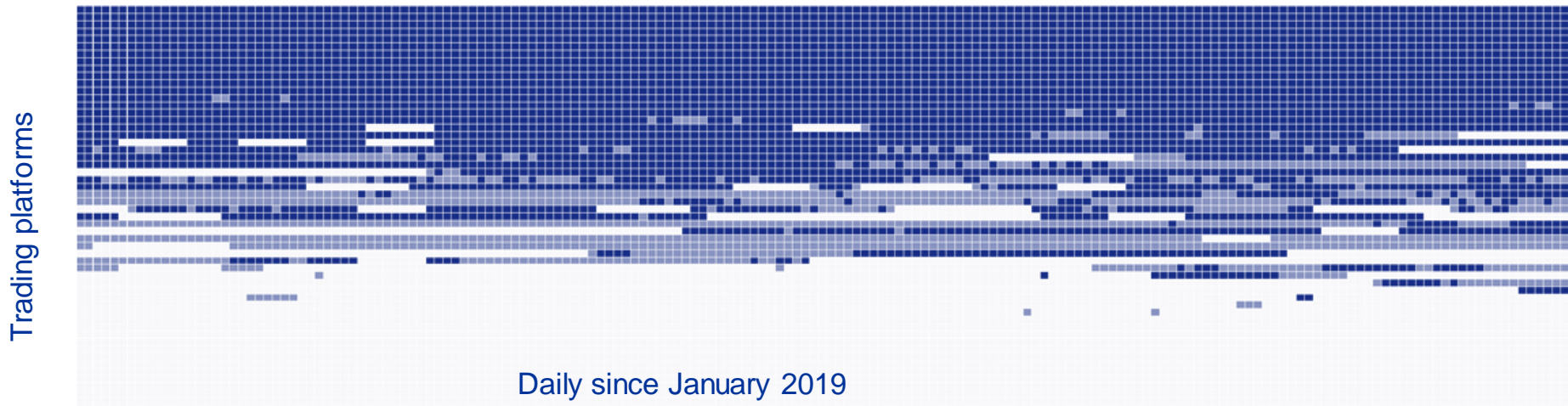
Input:

- close price and trading volumes for crypto-asset pairs; 1 week moving window of hourly data

6. Enhancing data quality – isolation forest

Isolation forest

- method to detect anomalies based on the concept of isolation, which does not employ any distance or density measure
- economic processing time, robust against masking and swamping
- works well in high dimensional problems and irrelevant attributes



Input:

- OHLC prices and daily changes in trading volumes for crypto-asset pairs; 1 week moving window of hourly data;
- trading platform features: centralised/decentralised, country of incorporation, fees/no-fees, availability of trading book, web searches results and web traffic*, sentiment based on news*

6. Enhancing data quality – dependency network analysis

Cyclic nowcasting/prediction analysis

Input

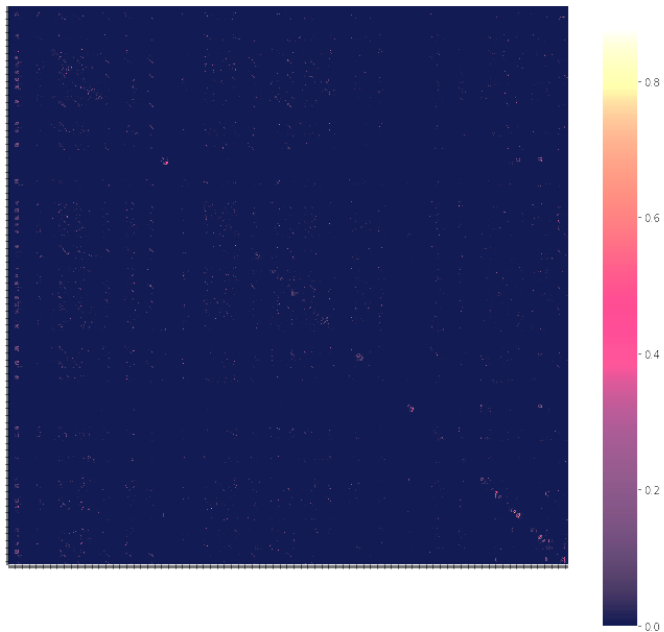
- Pricing and trading data, trading platforms metadata*

Each attribute is used as a target, iteratively.
All the others are then used as predictors.

XGBoost is used for each regression.
N attributes == N regressions

Result -> Feature influence

Influence/Dependence Heatmap between all features

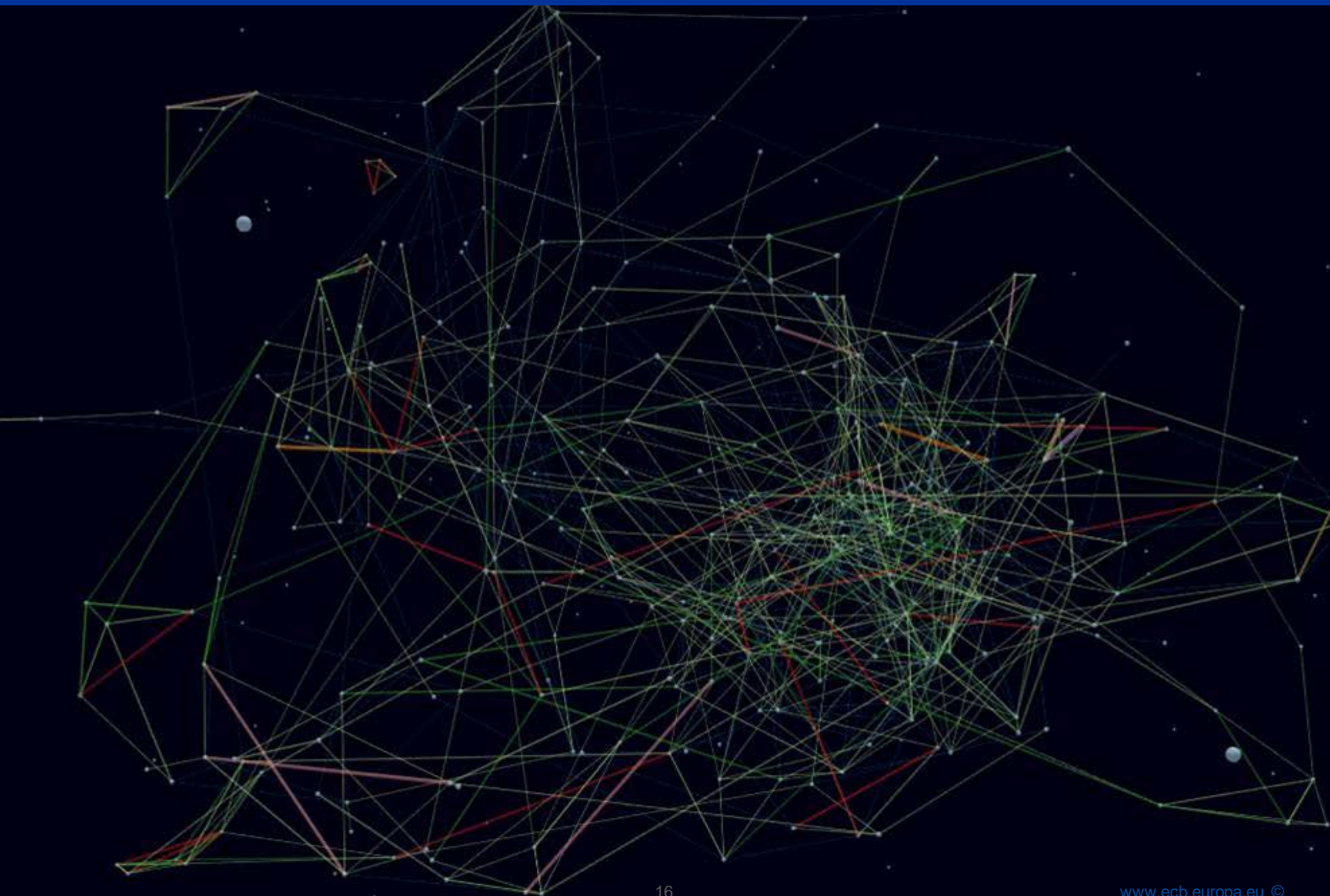


Dependency network

- Interactive 3D force graph
- Interactive 2D exchange aggregation

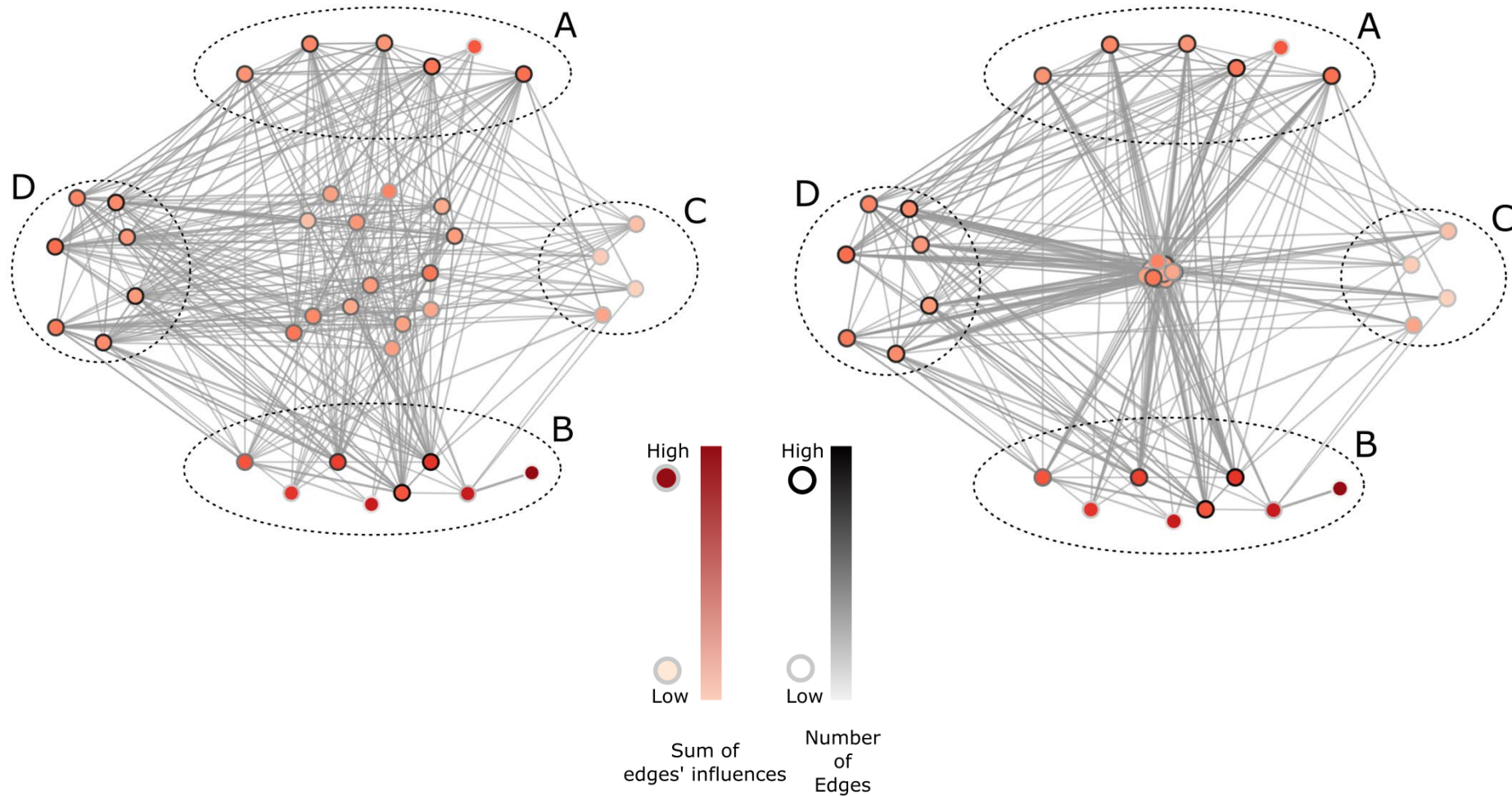
*(see next slides –
examples for BTC/EUR pairs)*

Dependency Network between all features (Interactive 3D force graph)



6. Enhancing data quality – dependency network analysis

Dependency Network between all features (Interactive 2D exchange aggregation)



Crypto trading platforms non-directed relationships on the BTC-EUR pair

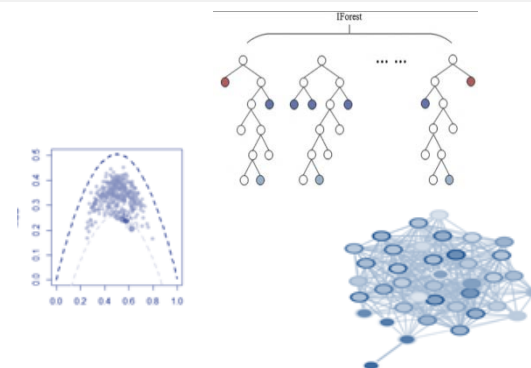
7. Conclusions

In the light of the implications they might have for the stability and efficiency of the financial system and the economy, crypto-assets **warrant continuous monitoring.**



To this end, the ECB has set up a **dataset and indicators** based on publicly available aggregated data complemented with other data from some commercial sources using API and big data technologies.

The ECB has investigated and implemented* **data quality enhancements algorithms** based on the functional data analysis, isolation forest as well as dependency network analysis in order to arrive at the fit for purpose dataset and indicators.



Future work will focus on **developing and enhancing indicators and closing data gaps** as new data and analytical needs may well arise with further advancements in crypto-assets and related innovation.

Thank you for your attention.