M2 DSC - Université Jean Monnet

Bitcoin Analysis

Data Mining for Big Data

Idriss BENGUEZZOU
Mohammed BOUTOUIZERA
Ghilas MEZIANE
Mahmoud NASHAR

02/02/2024

Table des matières

1 Introduction

Bitcoin has changed how digital transactions work by being decentralized and using blockchain technology. This project looks at the connections between Machine Learning, Data Mining, and Network Analysis to find insights in the Bitcoin network. As a data science team in Information Technology, our goal is to provide a detailed analysis for an organization interested in blockchain and Bitcoin. This report outlines our study, covering two main tasks that contribute to a full understanding of the Bitcoin ecosystem.

Our analysis plan in this report focuses on predicting Bitcoin prices, addressing the challenge of volatility, and understanding price trends. We also explore the community structure within Bitcoin transaction networks, examining how it changes over time and its potential links to Bitcoin price changes.

2 Data Description

The dataset under examination spans a period of two and a half years, from January 1, 2015, to June 30, 2017. It encompasses two primary types of data : Time Series and Transaction Networks, providing a comprehensive view of the Bitcoin ecosystem during this timeframe.

2.1 Time Series Data

The Time Series data is a crucial component of our study, offering insights into various aspects of the Bitcoin network, including Bitcoin price dynamics, mining activities, and detailed transaction information. Each file within this category is meticulously categorized by the corresponding year, with a particular emphasis on the year 2015.

In the context of our time series data analysis, we leverage three distinct CSV files namely, "External," "by_actor," and "global" each offering a unique perspective on the intricate dynamics of the Bitcoin ecosystem.

The "External" file serves as a gateway to understanding transactions occurring beyond the Bitcoin blockchain. It presents daily records of the currency's price and hash rate, offering a glimpse into the computational power driving the blockchain network.

Moving to the "by_actor" file, it delves into the activities of the top 100 significant actors within the Bitcoin realm. This file provides a detailed breakdown of transactions, imports, and associated statistics, shedding light on the behavior of influential actors shaping the landscape.

Lastly, the "global" file aggregates data from the Bitcoin blockchain, encompassing transaction specifics and actor behaviors. Encompassing details on transactions, payments, fees, mining activities, and self-spent sums, this file stands as a comprehensive resource, providing valuable insights into the overall dynamics governing the Bitcoin network. Together, these three files contribute to a holistic understanding of the temporal evolution and intricacies within the Bitcoin ecosystem.

2.2 Transaction Networks

The Transaction Networks section comprises daily transaction network files that encapsulate exchanges between major participants in the Bitcoin network. Each file within this category delineates key attributes, such as source, target, transaction value, and the count of transactions. Table [1] provides a sample of the data :

Table 1 – Sample of Transaction Network Data

Source	Target	Value	Number of Transactions
ePay.info_CoinJoinMess	CloudBet.com	3,519,173	1
Cex.io	1956	8,491,196	3
157228	C-Cex.com	10,833,480,021	1

These records form the foundation for our analysis of the community structure within the Bitcoin transaction networks. The attributes such as "Source" and "Target" denote the participants involved, "Value" represents the transaction value, and "Number of Transactions" indicates the count of transactions between the specified participants. This detailed information is crucial for understanding the dynamics and relationships within the Bitcoin transaction network.

3 Price Prediction Analysis

In this section, we embark on the challenging task of predicting the price dynamics of Bitcoin (BTC). Our approach involves two distinct strategies aimed at forecasting the future price movements of this cryptocurrency. The first method employs a classification framework, seeking to predict whether the price will increase or decrease based on the information from the preceding 6 days. The second method is focusing on predicting the actual numerical value of Bitcoin's price at a given time.

3.1 Price variation prediction

3.1.1 Overview

The primary focus is on conducting a comprehensive study of three CSV files within the project. Our goal is to extract essential and pertinent information from these files to develop a learning model capable of predicting whether Bitcoin prices will rise or fall based on weekly data.

The task involves thorough investigations to construct an intelligent model. This model, after being trained on historical time series data from a previous year, is expected to yield accurate predictions regarding the direction of Bitcoin prices on the seventh day of the week.
Importantly, these predictions will be based solely on the information available for the first six days of the week.

So, we have these three following CSV files for this task :
- The CSV file labeled "External" furnishes us with details regarding transactions occurring beyond the confines of the Bitcoin blockchain. This file supplies us with daily information on the price of the currency, as well as the hash rate, serving as a metric for the computational power within the blockchain network.
- The CSV file labeled "by_actor" offers insights into the activities of the top 100 significant actors within the Bitcoin ecosystem. It provides a comprehensive breakdown of

transactions, imports, and statistics associated with each transaction conducted by these influential actors.
— The CSV file labeled "global" compiles aggregated Bitcoin blockchain data, featuring transaction details and actor behavior. It includes information on transactions, payments, fees, mining activities, and self-spent sums. The file provides valuable insights into the overall dynamics of the Bitcoin network.

3.1.2 Feature engineering

First of all, we introduced an extra column labeled "variation" to the CSV file "external" This step was necessary for executing the initial procedure, which involves assessing how prices fluctuated according to the values provided in the sheet. Consequently, by comparing the current day's price with the previous day's, we could discern whether the value of the Bitcoin had decreased. Subsequently, we assigned -1 or +1 in the "variation" column, indicating a decrease or increase in price, respectively, for the following day.

To enhance the data set, we eliminated non-essential features across three CSV files. Specifically, the "Year," "Month," and "Day" features were excluded, as they do not contribute significantly to the predictive task. Instead, emphasis was placed on retaining the "Week" and "Weekday" features, which play a crucial role in representing time series data.
Following this refinement, the "global" CSV file underwent further optimization by discarding the "xxxx" feature. This decision was made based on its lack of meaningful relevance and its inability to provide valuable insights for the prediction task.
In the case of the "by_actor" CSV file, the "identity" feature was also eliminated. This choice was informed by the understanding that, in the subsequent feature extraction phase, the identity of each actor would not contribute significantly to the prediction. Moreover, extracting additional features from the "identity" feature was deemed impractical.

After that we began by implementing straightforward solutions to facilitate a comprehensive grasp of the problem, gradually progressing towards more intricate analyses to achieve our desired objective. Initially, our focus was solely on the "external" CSV file, where we trained our classification model using the information extracted from this dataset.

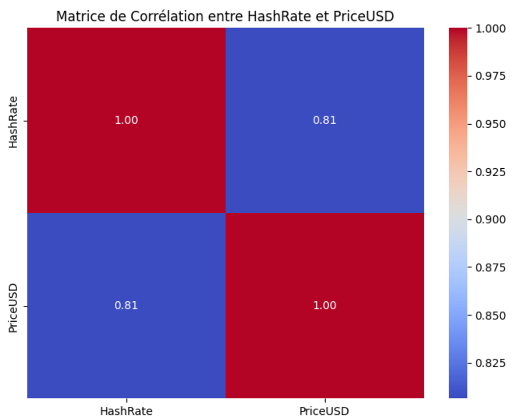

Figure 1 – Correlation PriceUSD and HashRate

We initiated this initial step based on the correlation matrix presented in the figure. As evident from the data, there is a notable high correlation between the price and the hash rate. So it was important for us to start our first experimentation to verify if we can obtain good results with the simple features.

The results we achieved showed bad accuracy for several models, ranging between 30% and 45%. As mentioned earlier, we initiated a more extensive research, considering the integration of new features to increase the capabilities of the classification models employed for predicting variations in Bitcoin prices.
In the second step, we combined the two CSV files, namely "external" and "global," to extract additional features aimed at enhancing the performance of our classification models in predicting BTC price variation.

However, our utilization extended beyond the previous two CSV files. We wanted to incorporate data from the third file, "by_actor," which can give us additional features not present in the other CSVs. Consequently, we conducted an analysis of this file, and we thought about the possibility of aggregating all activities occurring in a single day to create a comprehensive summary of a time series.

As we are aware, the "by_actor" file contains numerous features offering insights into the activities of each actor. To consolidate these activities on a time serie, we need to perform calculations that enable the aggregation of this information into a single field for each day. For instance, this could involve summing the number of transactions executed by the actors within a given day.
Following this operation, we obtain a single line for each day, where each column represents the sum of its respective cells. This entails consolidating the values of a particular feature into a singular cell for a given day.

Actually, we did not perform a simple summation. we applied for every feature :
- Standard Deviation (std) : We used the standard deviation to measure the variability or dispersion of values within each feature. A higher std may indicate greater diversity or volatility in the data.
- Minimum (min) and Maximum (max) : These operations were employed to identify the range of values for each feature. The minimum and maximum values offer an overview about the irregular actor's behaviors, helping us understand potential outliers or unique characteristics.
- Mean (average) : Calculating the mean allowed us to capture the central tendency of each feature. This provides a representative value that summarizes the average transaction behavior, helping in understanding the actor's typical financial activity.

With this operation, we can generate four new features for each existing column in the "by_actor" file corresponding to a day of the week.

At the conclusion of this step, we merge the three prepared CSV files by using the key (week, weekday). So we have now a single CSV file that contains all the features prepared for the classification models in the experimentation stage.

Finally, we did a final correlation study among all the features prepared and the price variation. However, the extensive correlation matrix did not reveal any significant correlations with the variation. Recognizing that the variation is inherently linked to the price, we performed an additional correlation study specifically between the 62 extracted features and the priceUSD. This analysis gave us promising correlation results with the other features.

Consequently, we decided to arrange the features in descending order based on their correlation score with the priceUSD, preparing them for use in the classification models.

3.1.3 Experimented models and Training

After the feature engineering phase, let's dive into the model selection and training parts.

3.1.3.1 Experimented models In this phase of the project, we focused on choosing machine learning models that can be efficient on Bitcoin price prediction tasks : Logistic Regression, Decision Tree, Random Forest, and Support Vector Classifier (SVC).

— **Logistic Regression :**Chosen for its simplicity and interpretability. It is well-suited for binary classification tasks, making it appropriate for predicting the variation in Bitcoin prices (up or down).
— **Decision Tree :** Selected for its ability to capture complex relationships in the data. Decision Trees are adept at handling non-linear patterns, which could be valuable in predicting Bitcoin price movements.
— **Random Forest :**Employed for its ensemble nature, combining multiple Decision Trees to improve predictive accuracy. Random Forest is known for its robustness and ability to handle noisy data.
— **Support Vector Classifier (SVC) :**Chosen for its effectiveness in handling high-dimensional data and its potential to capture intricate patterns in the Bitcoin price data.

3.1.3.2 Data Splitting Given the sequential nature of the data, a careful approach was taken to split the dataset. A split of 80% for training and 20% for testing was employed, respecting the temporal order of the data. This sequential split ensures that the models are trained on historical data and tested on more recent observations.

— **Avoidance of Shuffled Split :**Shuffled splitting or random cross-validation was avoided to maintain the time series order. This ensures that the models are exposed to past data before making predictions on future data points.
— **Balanced Splitting :**To enhance model performance, efforts were made to maintain class balance in both the training and testing datasets. This means ensuring a nearly equal number of instances for both upward (1) and downward (-1) price movements in each split. Additionally, to preserve the sequential nature of the data, the decision was made to avoid dividing the week. Consequently, the entire week was considered as a unit during the dataset splitting process.

3.1.3.3 Model Training and Testing Following the feature engineering phase and an in-depth study of feature correlations, we tried to identify the most influential features. To do so , our approach was to select the best features, starting from a set of two and gradually increasing the feature count up to 40.

— **Best Feature Iteration :**The training and testing procedure was executed iteratively, beginning with the top two features. Subsequently, additional features were incrementally incorporated into the models, assessing the impact on accuracy.
— **Sequential Evaluation :**Throughout this process, sequential testing was maintained, respecting the chronological order of the data. This approach ensured that each model's performance was assessed based on its ability to generalize to future time points.
— **Balanced Accuracy Consideration :**The evaluation of model performance consistently took into account balanced accuracy. This metric was crucial in mitigating the influence of class imbalance, providing a fair assessment of predictive power across varying feature sets.

This iterative approach not only allowed for the identification of the most influential features but also provided valuable insights into the relationship between feature count and predictive accuracy, ultimately guiding the selection of the optimal set of features for the final models.

3.1.4 Features selection

In the process of feature selection and model optimization, a systematic approach was adopted to determine the optimal feature set for each of the four chosen models : Logistic Regression, Decision Tree, Random Forest, and SVC. This involved comprehensive model evaluation with feature counts ranging from 2 to 40. Through iterative testing, starting with the best two features and incrementally adding more, each model's performance was meticulously recorded, creating a comprehensive record of accuracy scores. The synthesis of these evaluations facilitated the identification of the optimal feature count for each model. Ensemble analysis of accuracy scores allowed discernment of patterns, revealing the feature count at which each model achieved peak performance. Combining these results illuminated the feature count yielding the highest accuracy for each model, providing nuanced insights into their preferences. The iterative process also uncovered the importance of specific features in enhancing predictive accuracy. Armed with these insights, informed decisions were made regarding the final selection of features for each model. This methodical approach not only identified optimal feature sets but also provided valuable insights into the relationships between feature count and predictive accuracy, serving as a robust foundation for our final model configurations.

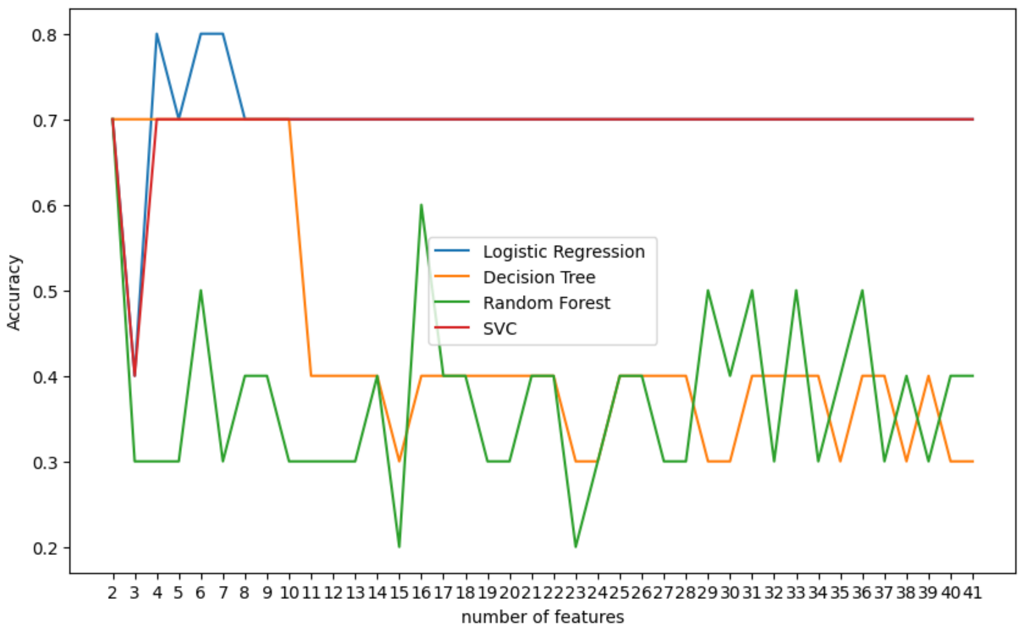

FIGURE 2 – Evolution of Accuracy with Number of Features for Models

Before the final model selection, as depicted in 2, it is apparent that a significant number of models achieved their peak accuracy when utilizing six features. This observation served as a foundational criterion for our ultimate feature selection process.

```
selected_features = ['weekday', 'week', 'PriceUSD', 'HashRate',
'nb_transactions', 'total_received']
```

3.1.5 Model Selection

Following the extensive testing of the four chosen models Logistic Regression, Decision Tree, Random Forest, and SVC. Each model was rigorously evaluated based on various performance metrics, including accuracy, precision, recall, and F1-score. A comparative analysis was conducted to assess how well each model adapted to the sequential nature of the data and exhibited robust predictive capabilities.
In making the final decision, several factors were considered, with a primary focus on accuracy, interpretability, and generalizability to unseen data. The accuracy of each model was a key factor in determining its effectiveness in predicting Bitcoin price variations. Logistic Regression, being a linear model, offered a distinct advantage in terms of interpretability. This characteristic was particularly valuable for understanding the factors influencing the model's predictions. Generalization to unseen data, a crucial aspect in time series prediction.
After a thorough comparison of the performance and accuracy of each model, Logistic Regression emerged as the preferred choice for the prediction task in Task 1.

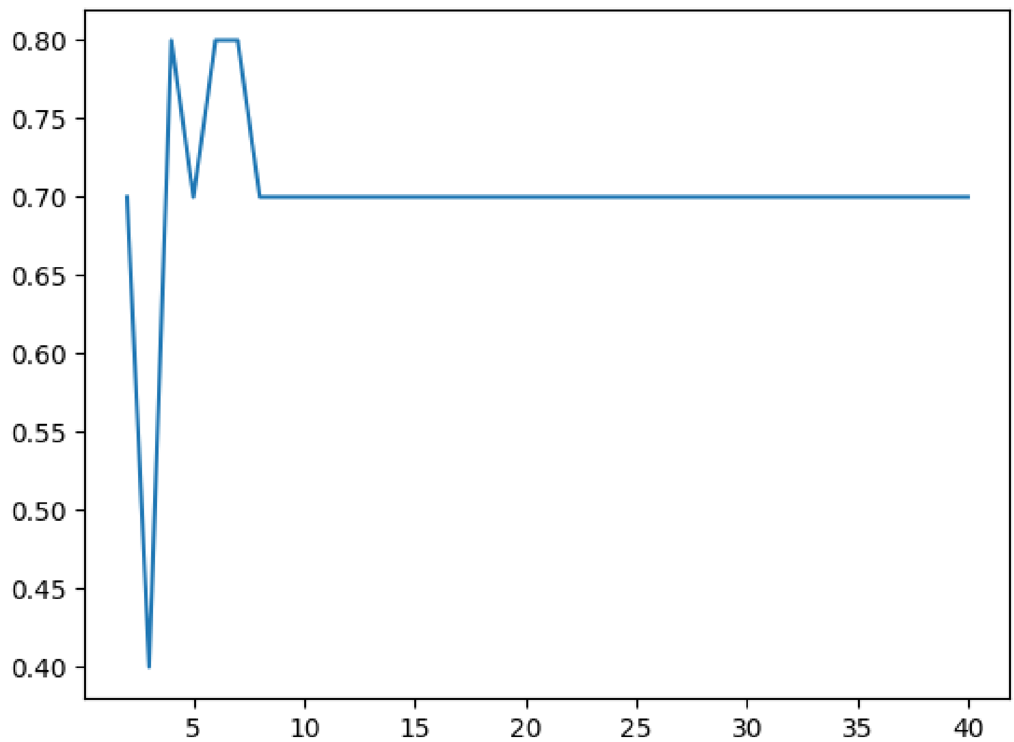

FIGURE 3 – Evolution of Accuracy with Number of Features for Logistic Regression

3.2 Price prediction

After the prediction of the variation, we wanted to go further by trying to experiment different machine learning techniques, models etc.. to generate a prediction of the Bitcoin's price value.

3.2.1 Exploratory Analysis

First of all we started by implementing an exploratory analysis notebook in order to :
- Gain a comprehensive understanding of the dataset's structure and content.
- Identify key features that may influence Bitcoin's value.
- Detect any anomalies or patterns within the data that could impact the accuracy of the predictive model.
- Formulate hypotheses about the relationships between different variables and Bitcoin's value.

In our initial exploratory analysis, we started by examining the presence of missing values across our datasets, specifically in the external, global, and by_actor dataframes. This check revealed that there were no missing values. Following this, we delved into visual explorations to understand the distribution and trends within our data. We first plotted a histogram of the Bitcoin price in USD, revealing the frequency distribution of prices over the dataset's timeframe.

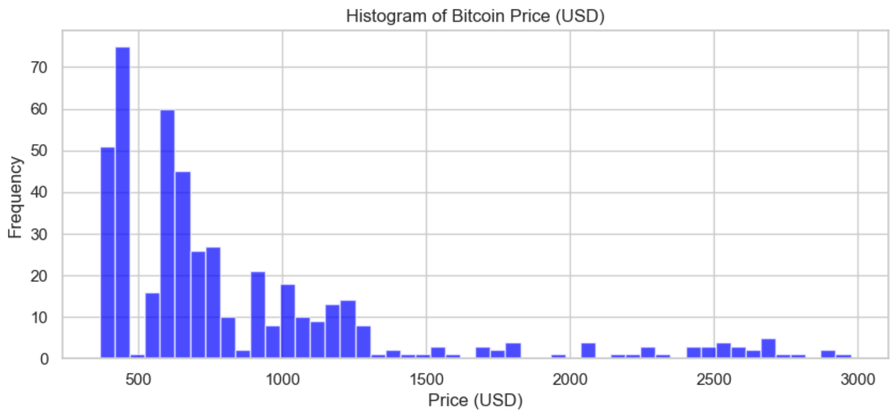

FIGURE 4 – Frequency by price histogram

Here we can see that the majority of Bitcoin prices are concentrated in the lower price range, as indicated by the tallest bar on the left side of the histogram. But, the distribution contains a long right tail, suggesting there have been periods when the Bitcoin price reached significantly higher values, although these instances are less frequent. The clustering of frequencies in the lower price ranges implies that Bitcoin has spent most of the observed time period at lower values, but the peaks at certain price intervals means for us that our prediction models should take into consideration these 'outliers' values.

To understand the temporal dynamics of Bitcoin prices, we also plotted the average weekly price. This line graph, marked by data points for each week, allowed us to observe fluctuations over time, offering a macro view of price movements on a weekly basis. The price of Bitcoin

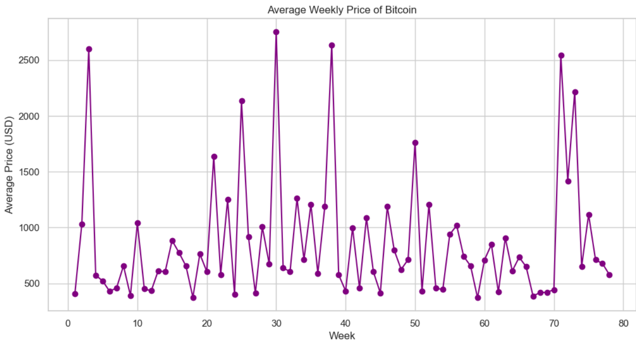

FIGURE 5 – Average weekly price

shows high volatility from week to week, with significant fluctuations in the average price. A cyclical pattern is not immediately apparent from this visualization, suggesting that the price changes are not following a simple, predictable pattern over the weeks. This is explainable by the fact that this plot was made on the task1/external file, and as specified in the project instructions, the weeks are not ordered. This dramatic rises and falls in average weekly prices should be well managed by our prediction models.

Finally, we explored the relationships between the Bitcoin price and other variables through correlation matrices. This analysis aimed to identify potential predictors that exhibit strong associations with Bitcoin's price, thus informing our feature selection for subsequent predictive modeling.

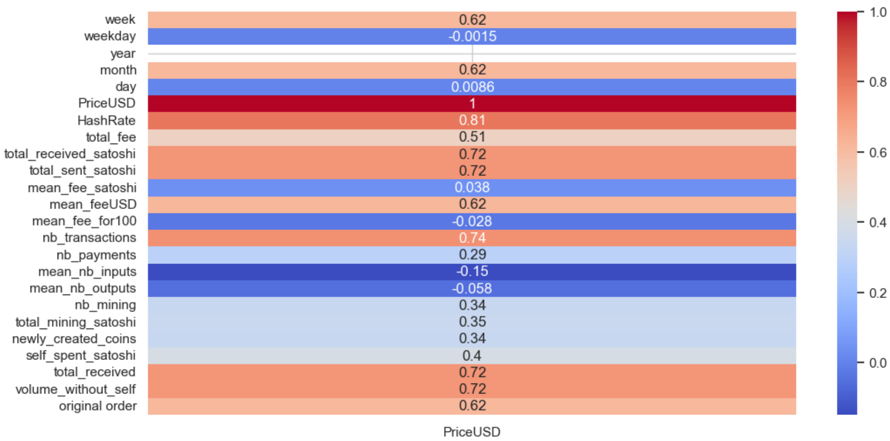

FIGURE 6 – Price Correlation

We can see that the high-correlated columns are : HashRate, nb transactions, total received,total sent satoshi,volume without self.

3.2.2 Deep learning Models

3.2.2.1 Metrics of evaluation In the evaluation of our predictive models, we decided to use a set of metrics to measure accuracy and reliability. These metrics provide insights into various aspects of the model's predictions, such as the average error, the consistency of the error, and the proportion of variance explained by the model. Below, we describe each metric used :

— **Mean Squared Error (MSE)** : Represents the average squared difference between the estimated values and the actual value. MSE is a measure of the quality of an estimator and is always non-negative. Lower MSE values indicate better fit to the data.

— **Root Mean Squared Error (RMSE)** : Is the square root of the MSE and serves as a measure of the deviation of the predicted values from the observed values. Like MSE, a lower RMSE is better.

— **Mean Absolute Error (MAE)** : Is the average absolute difference between the predicted values and the actual values, providing a linear score without considering direction. It is less sensitive to outliers compared to MSE and RMSE.

— **R-squared (R2)** : Indicates the proportion of variance in the dependent variable that can be explained by the independent variables in the model. It is a statistic that will give some information about the goodness of fit of a model. An R2 of 1 indicates that the regression predictions perfectly fit the data.

— **Mean Absolute Percentage Error (MAPE)** : Expresses accuracy as a percentage of the error, and it is a measure of prediction accuracy of a forecasting method in statistics. It is defined as the average absolute percent error for each time period minus actual values divided by actual values.

— **Variance Ratio** : Is used to measure how much of the variability in the data is accounted for by the model. A lower variance ratio suggests that the model is not overfitting and is generalizing well to new data.

3.2.2.2 Ensuring Reproducibility of Results In machine learning, particularly in the context of neural networks, the initialization of weights and the stochastic nature of the algorithms can lead to variability in results. And we noticed this by getting different score values when we did several runs on the same model.

To resolve this issue and ensure the reproducibility of our results, we took several steps to set the random seeds for all the libraries that we used, which include Numpy, Random, and TensorFlow. By setting the random seeds we ensured that the randomly initialized weights of the neural networks and the random splits of the data remain consistent across different runs.

Moreover, we used TensorFlow's deterministic operations by enabling operation determinism, which ensures that given the same input and initial conditions, the operations will produce the same output on every run. We also set environment variables that control the hashing behavior in Python to maintain consistency in the order of operations.

```
def reset_seeds():
    np.random.seed(SEED)
    random.seed(SEED)
    tf.random.set_seed(SEED)
    tf.keras.utils.set_random_seed(1)
    tf.config.experimental.enable_op_determinism()
    os.environ['TF_DETERMINISTIC_OPS'] = '1'
    os.environ['PYTHONHASHSEED'] = str(SEED)
```

FIGURE 7 – Resetting the seeds to ensure reproducibility

By taking these measures, we aim to achieve deterministic behavior in our experiments, allowing us to produce the same results each time the models are trained and evaluated, thereby ensuring the reliability and credibility of our findings.

3.2.2.3 Training Setup The initial phase of our training and experimentation was conducted using the `2015/external.csv` file, which contains the Bitcoin price data for the year 2015. Our preliminary observations indicated that the price during this period was relatively stable, predominantly ranging between 300 to 500 USD, which did not provide the volatility necessary to rigorously test our models' predictive capabilities.

To address this, we constructed an augmented dataset, `external_with_outliers_in_test.csv`, and `external_with_outliers_in_train.csv` by deliberately introducing synthetic outliers into the test set. This modified dataset was created to simulate significant price fluctuations, allowing us to observe and evaluate how our models respond to volatile data, which is characteristic of cryptocurrency markets.

Additionally, we utilized the `task1` dataset to further our evaluation by predicting the price on the 5th day based on the preceding sequence of days.

This process necessitated the creation of testing windows where sequences of 6 days (or 5 days for task1/external.csv) (t, t+1, ..., t+5) were allocated to the feature set X, and the corresponding 7th day(or the 6th day or task1/external.csv (t+6) was assigned to the target set Y.

The data was then partitioned into training and testing sets with an 80%-20% split. The training set consists of 80% of the data sequences, which is used to train the models, while the remaining 20% forms the testing set, which is used to evaluate the models' performance on unseen data. This approach ensures that our models are trained on a diverse range of scenarios, including both typical and atypical market behaviors, thereby enhancing the robustness of our predictive analysis.

3.2.2.4 SVR Support Vector Regression (SVR) is a type of Support Vector Machine (SVM) that is used for regression tasks. In our experiments, SVR was considered due to its ability to handle non-linear relationships and to provide robust predictions even in the presence of noise, which is a common feature of financial time series data such as Bitcoin prices.

To optimize the SVR model, we performed a grid search over a set of hyperparameters. The hyperparameters included the type of kernel function ('rbf', 'poly', 'sigmoid'), the penalty parameter 'C' that controls the trade-off between achieving a low error on the training data and minimizing the model complexity, the 'gamma' parameter which defines the influence of single training samples on the decision boundary, and the ‘epsilon‘ parameter which specifies the epsilon-tube within which no penalty is associated with predictions.

The grid search was carried out using a 5-fold cross-validation on the training set, with the mean squared error (MSE) as the scoring metric. The cross-validation process involved dividing the training data into five subsets, training the model on four subsets, and evaluating it on the fifth subset. This was repeated five times to ensure each subset was used as the evaluation set. The combination of hyperparameters that resulted in the lowest MSE was selected as the best model.

C	epsilon	gamma	kernel
10	0.01	0.1	rbf

Table 2 – Best parameters for SVR

Results

— **2015/external.csv** : features=['PriceUSD','HashRate'], timestep=6

MSE	**RMSE**	**MAE**	**R2**	**MAPE**	**Variance Ratio**
1333.62	36.5188	30.4252	0.589319	0.0761945	0.485582

Table 3 – Performance Metrics of the Model

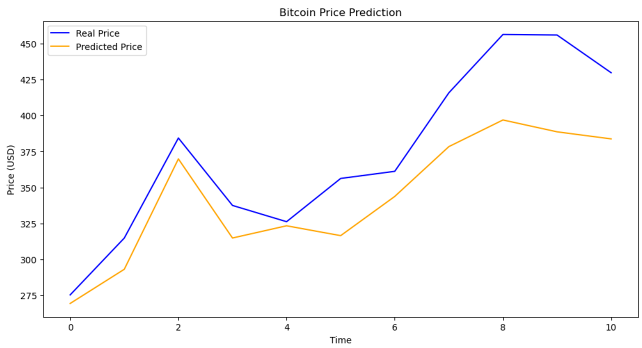

Figure 8 – Prediction results for SVR trained and tested on 2015/external.csv

— **task1/external.csv** : features=['PriceUSD','HashRate'], timestep=5 Here each 5 days are taken into consideration to predict the 6th.

MSE	**RMSE**	**MAE**	**R2**	**MAPE**	**Variance Ratio**
704.252	26.5377	19.1365	0.998256	0.022263	0.057446

Table 4 – Performance Metrics of the Model

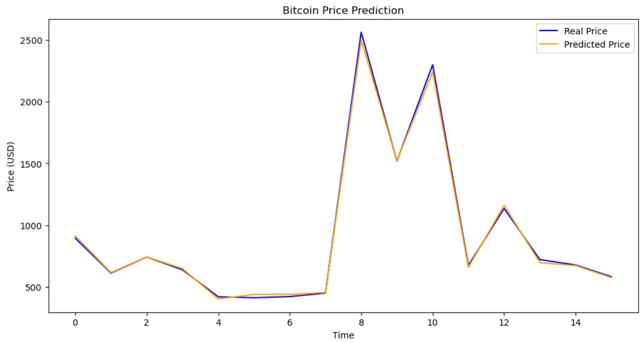

FIGURE 9 – Prediction results for SVR trained and tested on task1/external.csv

3.2.2.5 GRU Gated Recurrent Unit (GRU) is a variant of the LSTM network which has similar features, but with a simpler structure. GRUs are known for their efficiency in sequence prediction tasks, which makes them suitable for time-series forecasting such as predicting Bitcoin prices.

Architecture : The GRU model in our experiments is designed with an architecture that consists of two GRU layers followed by a dropout layer and a dense layer for output. The first GRU layer returns the full sequence to the next GRU layer, allowing the model to capture temporal relationships at different scales. The dropout layer is included to prevent overfitting, ensuring that the model generalizes well to unseen data.

Results :

— **2015/external.csv** : features=['PriceUSD','HashRate'], timestep=6

optimizer	loss function	epochs	batch size
Adam	mean absolute error	100	8

TABLE 5 – Training parameters

MSE	RMSE	MAE	R2	MAPE	Variance Ratio
186.239	13.6469	11.0654	0.942649	0.0296085	0.0729214

TABLE 6 – Performance Metrics of the GRU Model

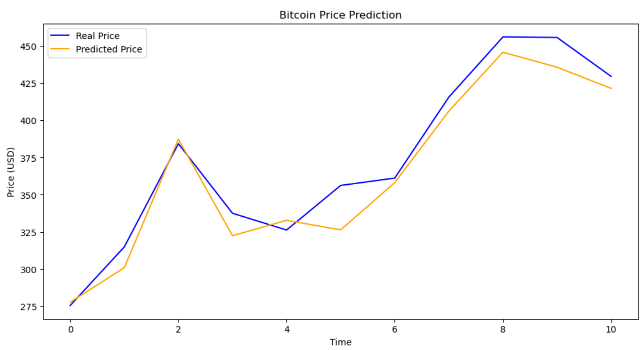

FIGURE 10 – Prediction results for GRU trained and tested on task1/external.cs

— **task1/external.csv** : features=['PriceUSD','HashRate'], timestep=5

optimizer	loss function	epochs	batch size
Adam	mean absolute error	100	8

TABLE 7 – Training parameters

MSE	RMSE	MAE	R2	MAPE	Variance Ratio
7977.35	89.316	58.0935	0.980243	0.0601669	0.212357

TABLE 8 – Performance Metrics of the GRU Model

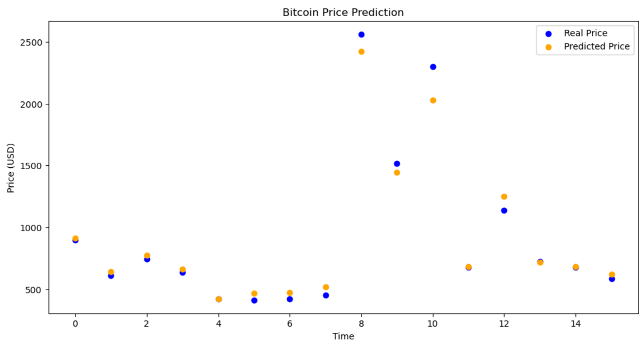

FIGURE 11 – Predictions results of GRU model on task1/external

We can observe that for the "Task 1 External" dataset, the model has a relatively high MAE and a higher R^2 score, suggesting the model's predictions are less accurate but explaining more variance. In contrast, for the "2015 External" dataset, the MAE is significantly lower, but the R^2 score is less higher, indicating better predictive accuracy. That's completely explainable when we look at the 2015 data that doesn't contain 'outliers'.

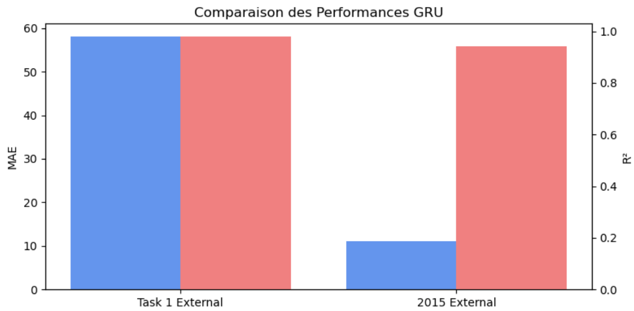

FIGURE 12 – Performance comparison GRU - task1 vs 2015

3.2.2.6 LSTM Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) capable of learning long-term dependencies, particularly useful for sequence prediction problems. In our study, we employed LSTMs due to their proven effectiveness in handling time-series data, where the prediction of future values often depends on understanding and processing the past information accurately. We conducted a series of experiments with the LSTM model to predict Bitcoin prices based on historical data.

Architecture : The model consists of two LSTM layers, each with 256 units. The use of multiple LSTM layers is intended to increase the model's capacity to learn complex patterns in the data. The inclusion of a Dropout layer with a rate of 0.2 is to prevent overfitting by randomly dropping out units in the neural network during training, which helps the model generalize better to unseen data.

The model concludes with a Dense layer with a single unit to output the predicted price value. This structure allows the LSTM to map the input sequence data to a continuous value, making it suitable for regression tasks such as predicting the price of Bitcoin.

Results :

— **2015/external.csv** : features=['PriceUSD','HashRate'], timestep=6

optimizer	loss function	epochs	batch size
Adam	mean absolute error	100	8

TABLE 9 – Training parameters

MSE	RMSE	MAE	R2	MAPE	Variance Ratio
232.01	15.2319	12.7427	0.928554	0.0359043	0.111589

TABLE 10 – Performance Metrics of the LSTM Model

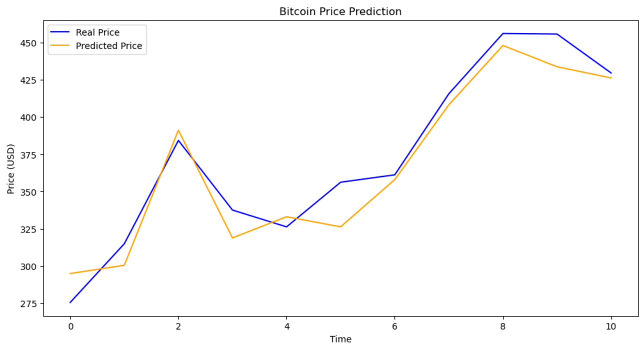

FIGURE 13 – Predictions results for LSTM on 2015/external.csv

— **task1/external.csv** : features=['PriceUSD','HashRate'], timestep=5

optimizer	loss function	epochs	batch size
Adam	mean absolute error	100	8

TABLE 11 – Training parameters

MSE	RMSE	MAE	R2	MAPE	Variance Ratio
809.805	28.4571	19.8018	0.997994	0.025248	0.0294225

TABLE 12 – Performance Metrics of the LSTM Model

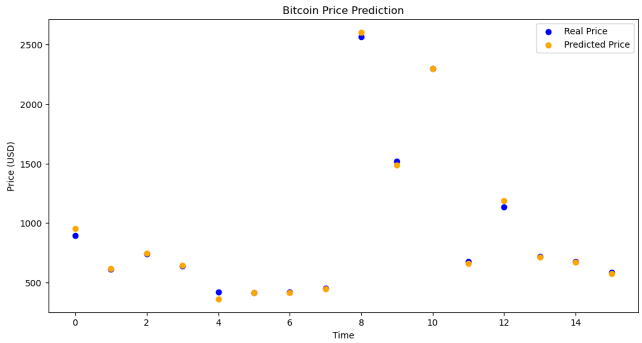

FIGURE 14 – Predictions Results for LSTM on task1/external.csv

For Task 1 External, the MAE is around 20, which is considerably acceptable for predictive modeling tasks. This suggests that on average, the model's predictions are within a 20$ range of the actual prices, indicating a relatively high level of accuracy in capturing the day-to-day variations in Bitcoin prices.

For 2015 External, the MAE is slightly lower, which implies that the model performs slightly better on this dataset. However, the model's R^2 score, which is a measure of the proportion of variance for a dependent variable that's explained by an independent variable or variables in a regression model, is below 1 for both datasets. This indicates that while the model has predictive power, there is still room for improvement in terms of explaining the full variability of the data.

Overall, the reduced difference in MAE between the two datasets demonstrates the model's consistent performance across different data conditions, with a particularly strong result for Task 1 External given the complexity of predicting highly volatile financial time series data like Bitcoin prices.

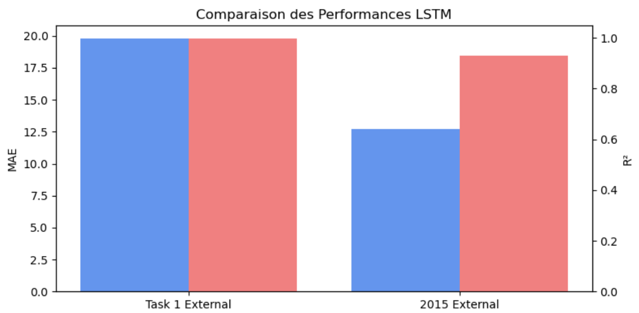

FIGURE 15 – LSTM Performance Comparison - 2015 vs task1

3.2.3 Model Selection

Finally, we opted to use our LSTM model to generate the predictions for the following reasons :

Consistent Performance : The LSTM model demonstrated a consistent performance across various tests, including handling datasets with and without artificially introduced outliers. This indicates a level of robustness in the model's ability to generalize from the training data to unseen data.

Handling Volatility : LSTM's inherent structure, with memory cells capable of maintaining information over time, proved effective in capturing the temporal dependencies and volatility in Bitcoin price movements, which are characteristic of financial time series data.

Reduced Mean Absolute Error : The LSTM model achieved a reduced MAE compared to other models, suggesting that it can provide predictions that are, on average, closer to the actual Bitcoin prices. This accuracy is crucial for the reliability of the predictions needed for the final submission.

Good R^2 Score : While not perfect, the LSTM model's R^2 score indicated a good fit to the data, suggesting that it could capture a significant proportion of the variance in Bitcoin prices.

Given these reasons, the LSTM model was selected as the most suitable for generating the final predictions due to its reliable performance, its ability to handle complex patterns in the data, and its overall predictive accuracy.

4 Network Analysis

In this section, we delve into the intricate structure and dynamics of Bitcoin's transaction networks. Our exploration focuses on daily exchanges among major participants, revealing key attributes such as source, target, transaction value, and count. Through network analysis methodologies, we seek to unveil patterns and correlations, shedding light on the evolution and community structures within Bitcoin's transactional landscape.

Our subsequent focus involved a meticulous analysis of the community structure within the Bitcoin transaction networks. The overarching goal was to gain a profound understanding of the temporal dynamics and homogeneity of these communities.

We endeavored to answer fundamental questions :

— Is the community structure stable over time, or does it undergo changes ?
— How homogeneous are the identified communities ?

Additionally, we explored the identification of key actors within each community, aiming to discern influential figures exerting significant influence over their respective communities.

4.1 Community Detection

Our primary objective was to identify communities within the network. The '*Network*' class in our experiment implements both the **Louvain** method and the **Girvan-Newman** method for community detection. Utilizing these methods, we sought to discern the inherent groupings within the Bitcoin transaction networks.

4.1.1 Community Detection : Louvain Method

Our exploration into community detection utilized the **Louvain** method. This method, known for its ability to efficiently handle large-scale networks, making it suitable for the complexity inherent in Bitcoin transaction networks.

The louvain algorithm requires the input graph to be undirected. The resulting communities represent major participants with shared transactional patterns.

This method identified a total of **six communities** within the Bitcoin transaction networks. Table [14] provides a summary of the identified communities, including the number of participants, total transaction volume, and the number of transactions.

Table 13 – Summary of Louvain Community Characteristics

Community ID	Volume	Nb Transactions	Nb Unique Transactions	Nb Actors	Avg Transactions/Actor	Avg Unique Transactions/Actor
1	468,832,339,336,904,756	30,327,386	3,542,608	3,481	8,712.26	1,017.7
4	71,793,322,779,010,889	1,920,300	249,541	589	3,260.27	423.669
6	38,036,926,674,557,446	3,931,898	1,403,105	4,685	839.253	299.489
2	2,539,459,537,692,511	886,701	171,234	622	1,425.56	275.296
5	1,663,252,660,176,170	582,544	501,285	549	1,061.1	913.087
3	26,017,838,398,784	17,593	8,961	49	359.041	182.878
Average	9.71486e+16	6.27774e+06	979,456	1,662.5	2,609.58	518.686

Checkout the **Community Detection Louvain** 27 section in the **Annexes** for more details.

Upon examining the identified communities, several notable observations emerge :

Community 1 stands out as the largest and most active community, with a total transaction volume of 468,832,339,336,904,756 and a total of 30,327,386 transactions. It also has the highest number of unique transactions, with 3,542,608. This community consists of 3,481 actors, resulting in an average of 8,712.26 transactions per actor and 1,017.7 unique transactions per actor which is two times the average of the other communities. However, we still don't know if the activity of this community is due to a single actor or multiple actors.

Community 4 has a significantly smaller transaction volume and number of transactions compared to Community 1. It consists of 589 actors and has an average of 3,260.27 transactions per actor and 423.669 unique transactions per actor.

Community 6, regroup 4,685 actors and make it the largest in terms of size. This community has a transaction volume of 38,036,926,674,557,446 and a total of 3,931,898 transactions. It has 1,403,105 unique transactions and 4,685 actors. On average, each actor in this community has 839.253 transactions and 299.489 unique transactions.

The remaining communities (2, 5, and 3) have even smaller transaction volumes, numbers of transactions, and numbers of actors compared to the previous communities. Further analysis will delve into the temporal dynamics and homogeneity of these communities, providing a more nuanced understanding of their evolution and significance within the broader network structure.

4.1.2 Community Detection : Girvan-Newman Method

To complement our analysis, we also utilized the **Girvan-Newman** method for community detection. This method is known for its ability to identify communities of varying sizes, making it suitable for the complexity inherent in Bitcoin transaction networks.

The Girvan-Newman use a hierarchical approach to identify communities. It starts by removing the edges with the highest betweenness centrality, and then it iteratively repeats this process until all the edges are removed. The resulting communities represent major participants with shared transactional patterns.

This method identified a total of **17 communities** within the Bitcoin transaction networks. Table [**??**] provides a summary of the identified communities, including the number of participants, total transaction volume, and the number of transactions.

Table 14 – Summary of Girvan-Newman Community Characteristics

Community ID	Volume	Nb Transactions	Nb Unique Transactions	Nb Actors	Avg Transactions/Actor	Avg Unique Transactions/Actor
Community 1	5.729e+17	36,296,002	5,249,339	8,861	4,096.15	592.409
Community 2	5.077e+15	957,808	569,875	779	1,229.54	731.547
Community 6	4.698e+15	48,320	15,313	36	1,342.22	425.361
Community 5	1.229e+14	46,724	9,261	51	916.157	181.588
Community 4	3.136e+13	85,037	6,879	55	1,546.13	125.073
Community 3	2.393e+13	25,285	19,411	139	181.906	139.647
Community 10	2.368e+13	2,293	2,567	6	382.167	427.833
Community 7	1.644e+13	4,484	1,221	13	344.923	93.9231
Community 11	3.418e+12	16,432	1,042	5	3,286.4	208.4
Community 13	2.084e+12	180,310	214	2	90,155	107
Community 14	3.335e+11	110	79	2	55	39.5
Community 12	2.631e+11	120	31	3	40	10.3333
Community 17	1.854e+11	139	27	2	69.5	13.5
Community 8	1.566e+11	1,174	923	10	117.4	92.3
Community 9	1.269e+11	289	197	7	41.2857	28.1429
Community 16	5.736e+10	1,547	178	2	773.5	89
Community 15	2.681e+10	348	177	2	174	88.5
Average	3.42877e+16	2.21567e+06	345,690	586.765	6,161.84	199.65

The Girvan-Newman method seems to have created more communities than the Louvain method. But this algorithm created one community that is much bigger than the others, the Community 1. This community has a transaction volume of 5.729e+17 and a total of 36,296,002 transactions. It has 5,249,339 unique transactions and 8,861 actors. On average, each actor in this community has 4,096.15 transactions and 592.409 unique transactions which is almost 6 times the average of the other communities.

The Community 2 is the second biggest community in terms of transaction volume, with a transaction volume of 5.077e+15 and a total of 957,808 transactions. This community is way smaller than the Community 1, it has a total of 779 actors and 569,875 unique transactions. On average, each actor in this community has 1,229.54 transactions and 731.547 unique transactions.

Except for the top four communities (1, 2, 6, 5), the other communities are relatively smaller in terms of transaction volume. However, we can notice that even with few members, some communities have a high volume of transactions, for example the Community 6 is on par with the Community 2 in terms of transaction volume, with a transaction volume of 4.698e+15 and a total of 48,320 transactions. Even though this community has only 36 actors and 15,313 unique transactions. On average, each actor in this community has 1,342.22 transactions and 425.361 unique transactions.

4.2 Evolution of the communities

The community structure within the Bitcoin transaction networks is not static. It undergoes changes over time as actors participate in different transactions. This section explores the temporal dynamics of the communities, seeking to discern the evolution of the community structure over time.

4.2.1 Community Evolution : Louvain Method

Our goal was to examine the evolution of the communities over time, seeking to discern the temporal dynamics of the community structure. We focused on the evolution of the six communities identified by the Louvain method, analyzing the changes in their transaction volumes over time.

The main purpose of this analysis was to determine whether the communities transitioned between different states over time or remained stable. We also sought to identify the communities activity were increasing or decreasing at the same time, if there was a kind of trend or pattern in the evolution of the communities.

So we plotted the evolution of the transaction volume of each community over time. The activity of the communities varies over time, with some communities experiencing a significant increase in activity, while others seems at first to remain stable.

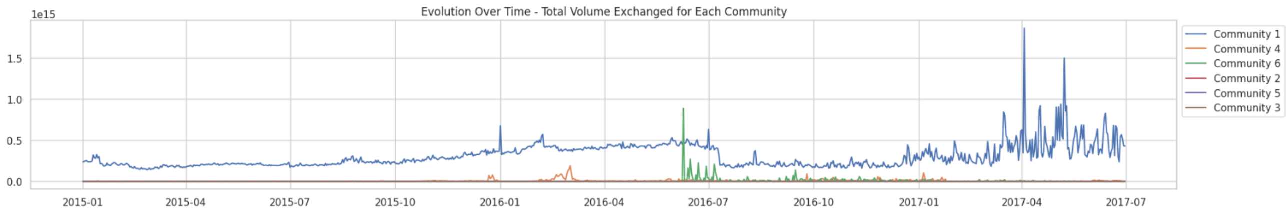

Figure 16 – Evolution over time - Total Volume exchanged by community

To have a better understanding of the evolution of the communities, we normalized the transaction volumes of each community by dividing them by the total transaction volume of the network at each time step, enable us to compare the evolution of the communities over time.

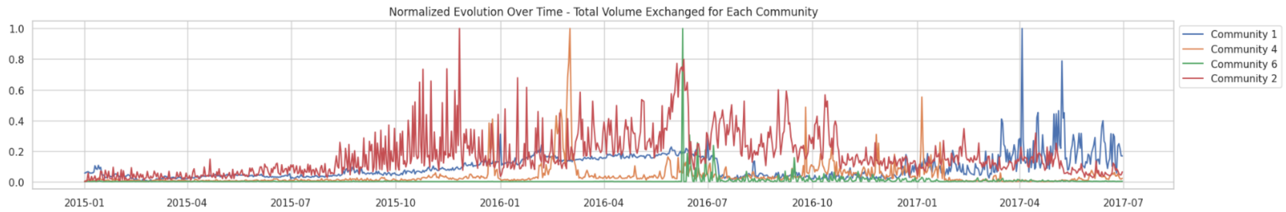

Figure 17 – Evolution over time - Normalized Total Volume exchanged by community

We can see that there is no kind of stability in the transaction volumes of the communities over time. We also notice that the communities are not active at the same time, for instance, Community 1, 4, 6 have a peak of activity at different periods of time. But at the same time we can notice that the Community 2 and 4 seems to overlap in their activity, arround the July 2016.

To refine our analysis, we applied a moving average to the normalized transaction volumes of each community, seeking to discern the underlying trends in the evolution of the communities over time.

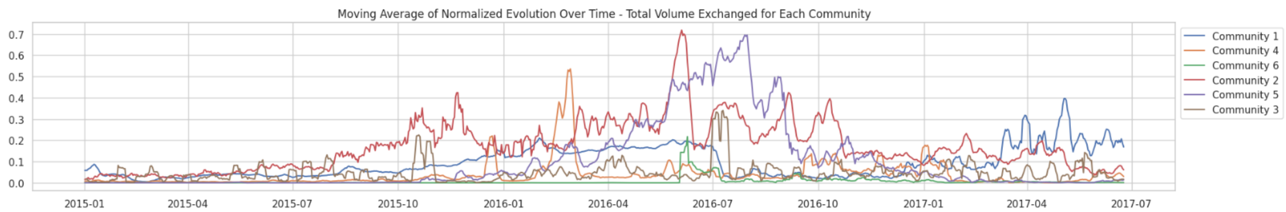

Figure 18 – Evolution over time - Moving Average of Normalized Total Volume exchanged by community

We can now start to see some kind of trend between the communities, for example, between 2016-04 and 2016-10, the transaction volumes of most communities are increasing significantly.

So we might think that the communities transaction volumes are correlated, wich means that at some period of time, the communities increase or decrease their activity. To confirm this hypothesis, we computed the correlation between the transaction volumes of the communities [19].

In conclusion, we can say that the Community 1 and 4 are the most active communities, and their activity is correlated over time. One hypothesis we can do now is that if we find a

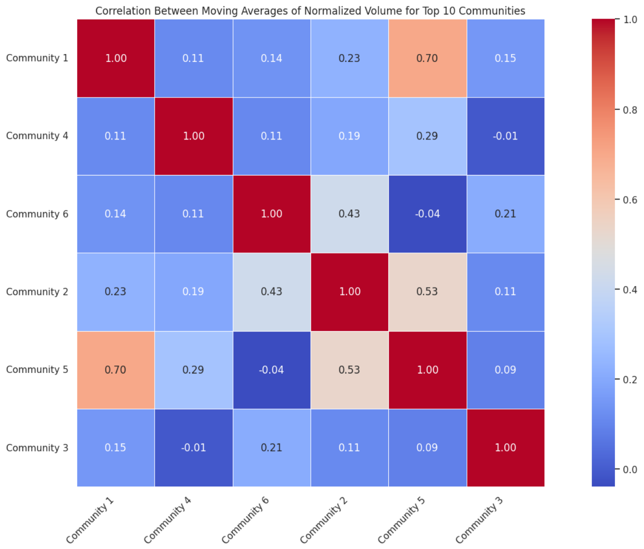

FIGURE 19 – Correlation Between Moving Average of Normalized Transaction Volumes

We can see that the transaction volumes of some communities are highly correlated, for example, the transaction volumes of Community 1 and 5 are highly correlated, with a correlation coefficient of 0.70. So some communities are correlated over time, which means that there is a high probability that the activity of these communities is linked in some way.

The transaction volumes of Community 2 and 5 are also correlated, with a correlation coefficient of 0.53. Understanding the correlation between the communities will help us to understand the evolution of the communities over time.

correlation between the activity of these communities and the price of Bitcoin, then we could be able to predict the trend of the price of Bitcoin.

4.2.2 Community Evolution : Girvan-Newman Method

As for the Louvain method, we plotted the evolution of the transaction volume of each community over time. Because of the difference in scale between the Community 1 and the other communities, we decided to directly plot the normalized transaction volumes of each community.

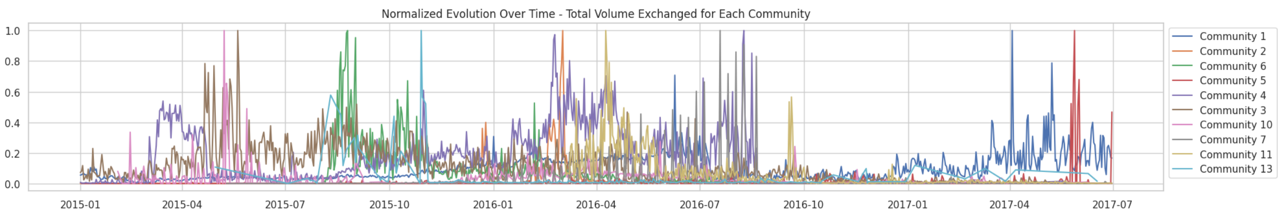

FIGURE 20 – Evolution over time - Normalized Total Volume exchanged by community

This is still quite hard to conclude anything from this but we can observe that there is some period of time where the communities are more active than others. For example, between 2016-10 and 2017-04 it seems that there is really few activity on the network. Let's try to apply a moving average to the normalized transaction volumes of each community, to see if we can find some trends.

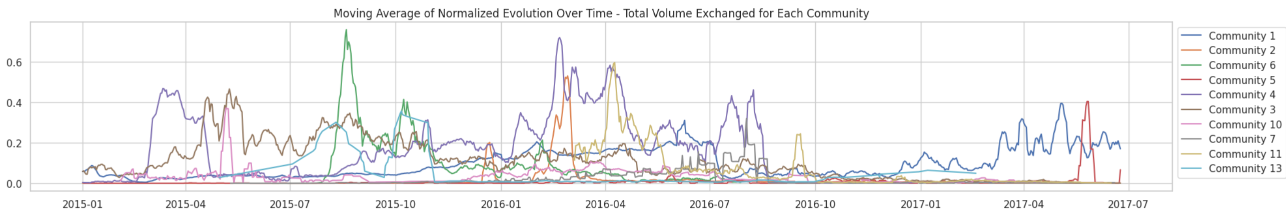

FIGURE 21 – Evolution over time - Moving Average of Normalized Total Volume exchanged by community

As for the Louvain method, we can now start to see some kind of trend between the communities, it is still hard to see it but it seems like most of the communities are active arround 2016-01 and 2016-10.

We will now try to find a correlation between the transaction volumes of the communities, to see if the communities are correlated over time.

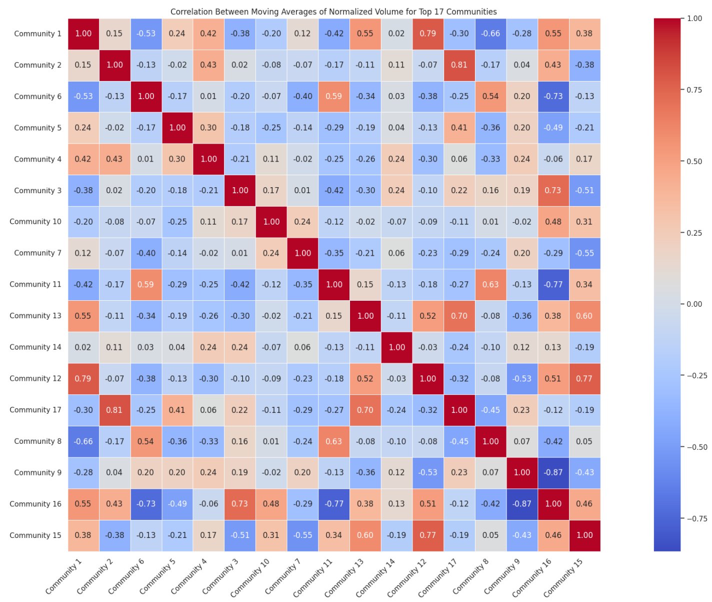

FIGURE 22 – Correlation Between Moving Average of Normalized Transaction Volumes

We can see that the transaction volumes of some communities are highly correlated, for example, in comparison with the Louvain method, the communities are more intricated, for example, the transaction volumes of Community 1 is correlated with the transaction volumes of Community 16, 12, 13.
These correlation reveal that the communities activity are linked and start at the same times and end at the same time for, certain communities.

As for the Louvain method, we can see that the transaction volumes of some communities are highly correlated. Finding a correlation between these communities and the price of Bitcoin, could help us predict the trend of the price of Bitcoin.

The key factor that will make us choose between the Louvain method and the Girvan-Newman method at this point is how much correlated are the communities with the price of Bitcoin. But before that, we would like to understand the homogeneity of the communities, is there key actors in each community that influence the activity of the community ?

We will try to analyze the top actors in each community, and see if these actors are what we call "key actors" that influence the activity of the community.

4.3 Communities Key Actors

In this section, we will try to identify the key actors in each community, and see if these actors are what we call "key actors" that influence the activity of the community.

4.3.1 Communities Key Actors : Louvain Method

First of all, here is a summary of the top tens actors indepedently of the community they belong to [31]. As we can see the biggest actor **ePay.info_CoinJoinMess** belongs to the biggest community **1**. Moreover out of the top ten actors, five of them belong to the biggest community **1**.

Our objective is to identify the key actors in each community, and see if these actors are what we call "key actors" that influence the activity of the community. To do so, we will plot the moving average of the normalized transaction volumes of the top ten actors in each community and try to find some correlation between the activity of the actors and the activity of the community.

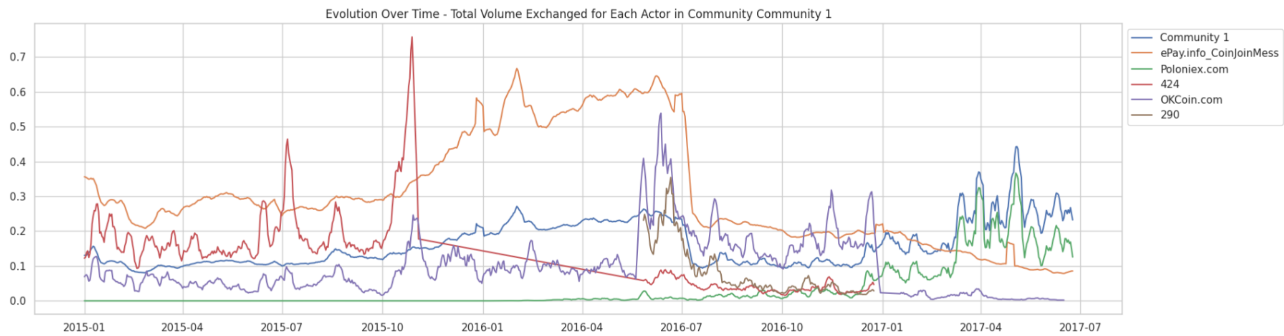

FIGURE 23 – Comparison of the Transaction Volumes of the Top Ten Actors in Community 1

As we can see it is difficult to find a correlation between the activity of the actors and the activity of the community, but we can still notice that the activity of the actor **ePay.info_CoinJoinMess** and **Poloniex.com** seems to be correlated with the activity of the community.

Morover, we can see that some actors are not active all the time, for example, the curve of the actor **ePay.info_CoinJoinMess**, is off between 2015-10 and 2016-07, suggesting that this actor was not active during this period of time. You can check the other communities in the **Annexes** [29].

As we can see it is difficult to find a trends between the activity of the actors and the activity of the community, by reading the plots. So we will try to find a correlation between the activity of the actors and the activity of the community, by computing the correlation between the moving average of the normalized transaction volumes of the actors and the moving average of the normalized transaction volumes of the community.

Of course we compute the correlation only for the actors that has high transaction volumes, because we want to find the key actors in each community. So we selected the top ten actors in each community, and computed their correlation with the community.

Poloniex.com is correlated with the activity of the community 1, with a correlation coefficient of 0.74. So these actors are what we call "key actors" that influence the activity of the community. If one of these Community is also correlated with the price of Bitcoin, then we could be able to

Table 15 – Correlation Between Moving Average of Normalized Transaction Volumes

Community ID	Actor ID	Correlation
1	Poloniex.com	0.74
2	Matbea.com	0.67
5	6048154	0.63

For our analysis we only considered the actors that have a correlation coefficient greater than 0.6. We can see that some actors are correlated with the activity of the community, for example, the actor

predict the trend of the price of Bitcoin by monitoring not only the activity of the community but also the activity of the key actors.

4.3.2 Communities Key Actors : Girvan-Newman Method

We also made a summary of the top tens actors indepedently of the community they belong to [32]. As we can see the biggest actors all belong to the biggest community **1**.

Our objective remain the same, we want to find the key actors in each community, and see if these actors are what we call "key actors" that influence the activity of the community. To do so, we will plot the moving average of the normalized transaction volumes of the top ten actors in each community and try to find some correlation between the activity of the actors and the activity of the community.

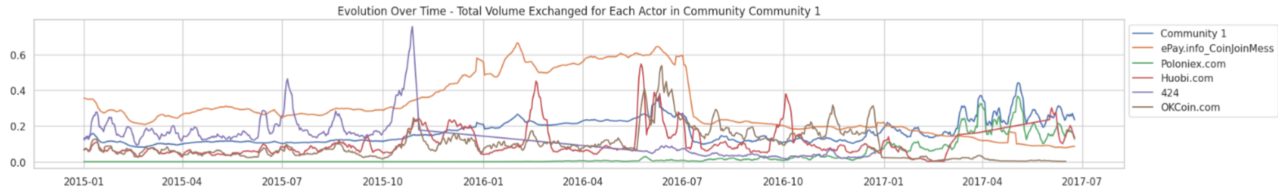

Figure 24 – Comparison of the Transaction Volumes of the Top Ten Actors in Community 1

As the biggest actor of the biggest community **1** is **ePay.info_CoinJoinMess**, and that the top tens actors all belong to the biggest community **1**, there is a high probability that key actor of the Community 1 remain the same.

To confirm this, we will compute the correlation between the moving average of the normalized transaction volumes of the actors and the moving average of the normalized transaction volumes of the community.

Table 16 – Correlation Between Moving Average of Normalized Transaction Volumes

Community ID	Actor ID	Correlation
1	Poloniex.com	0.71
2	396	0.70
11	53944833	0.93
13	74068096	0.99
12	979865	1.0
16	72183338	0.67
16	71554878	0.62

We still considered the actors that have a correlation coefficient greater than 0.6. We can see that **Poloniex.com** remained the key actor of the Community 1, with a correlation coefficient slightly lower than the Louvain method. The Community 12 has a correlation coefficient of 1.0, but could be explained by the fact that this community is quiet small, as there is only 3 actors. It is the same for the correlation of the Community 11 and 13.

The greater the number of actors inside a community the harder it is for an actor to influence the activity of the community. So having a correlation coefficient of 0.71 for the Community

1 with **Poloniex.com** is quite unexpected, because this community has 8,861 actors and the highest transaction volume.

Morover, having only one actor that influence the activity of the community is also unexpected, because we would expect that the activity of the community is influenced by multiple actors and not only one.

TABLE 17 – Actor Information for Poloniex.com

Name	**Community**	**Total Volume**	**Sent Volume**	**Received Volume**	**Total Transactions**	**Unique Transactions**
Poloniex.com	Community 1	58,719,270,696,167,512	55,802,900,064,914,770	2,916,370,631,252,742	1,410,104	88,038

We computed the ratio of the informations of **Poloniex.com** to understand the influence of this actor on the community.

As we are studying the volume sent, we can see that **Poloniex.com** sent 19.5% of the total volume of the community. Moreover, **Poloniex.com** received 1.01% of the total volume of the community. So we can see that **Poloniex.com** is more active in sending money than receiving it.

TABLE 18 – Ratios for Poloniex.com

	Ratio
Sended Volume	0.195013
Received Volume	0.0101707
Number of Transactions	0.0388501
Number of Unique Transactions	0.0167713

Now that we found our key actors, what remains now is to highlight any correlation between the activity of the communities and the price of Bitcoin to see if we can predict the trend of the price of Bitcoin by monitoring the activity of the communities and the key actors.

4.4 Communities and Bitcoin Price

In this section, we will try to find a correlation between the activity of the communities and the price of Bitcoin. We will plot the moving average of the normalized transaction volumes of the communities and the moving average of the normalized price of Bitcoin, and try to find some correlation between the activity of the communities and the price of Bitcoin.

4.4.1 Communities and Bitcoin Price : Louvain Method

We plotted curve to compare the behavior of the communities and the price of Bitcoin over time. The results are shown in Figure 33.

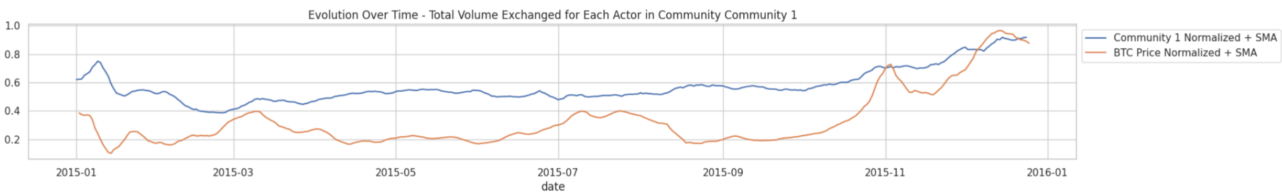

FIGURE 25 – Comparison of the Bitcoin Price and the Transaction Volumes of Community 1

As we can see, there seem to be a high correlation between the activity of the community 1 and the price of Bitcoin. The Community 1 is the biggest community in terms of transaction volume, and it is also the most active community.

Community ID	Correlation
1	0.71
4	0.49
6	-0.21
2	0.37
5	0.68
3	-0.06

Table 19 – Correlation BTC / Community

We can see that the activity of the community 1 is indeed highly correlated with the price of Bitcoin, with a correlation coefficient of 0.71. Morover, the activity of the community 5 is also correlated with the price of Bitcoin, with a correlation coefficient of 0.68.
The Community 4 is also correlated with the price of Bitcoin, with a correlation coefficient of 0.49, but the correlation is not as strong as the Community 1 and 5.

Here is a summary of the correlation between the activity of the communities and the price of Bitcoin.

So based on these results, we can say that the activity of the communities 1, 4 and 5 are correlated with the price of Bitcoin. So if we monitor the activity of these communities, we could be able to predict the trend of the price of Bitcoin, based on the activity of these communities. Morover, from our previous analysis, we know that the activity of the community 1 is correlated with the activity of the actor **Poloniex.com**, and the activity of the community 5 is correlated with the activity of the actor **6048154**.

4.4.2 Communities and Bitcoin Price : Girvan-Newman Method

We plotted curve to compare the behavior of the communities and the price of Bitcoin over time. The results are shown in Figure 34.

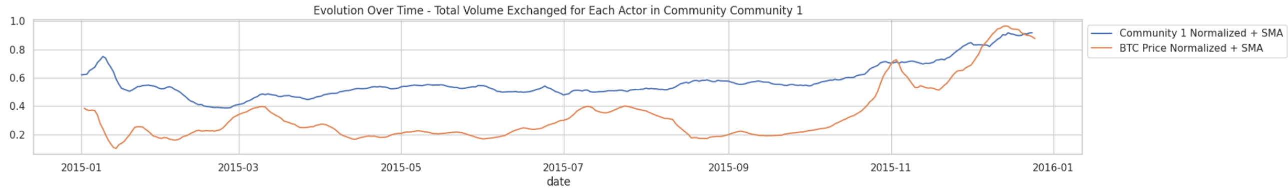

FIGURE 26 – Comparison of the Bitcoin Price and the Transaction Volumes of Community 1

As for the Louvain method, there seem to be a high correlation between the activity of the community 1 and the price of Bitcoin. This Community is the biggest community in terms of transaction volume, and we saw that **Poloniex.com** is the key actor of this community. So monitoring the activity of this actor and the activity of the community 1, could help us predict the trend of the price of Bitcoin.

Here is a summary of the correlation between the activity of the communities and the price of Bitcoin.

Community ID	Correlation
1	0.77
2	0.35
6	-0.25
5	-0.027
4	0.24
3	-0.22
10	0.09
7	0.22
11	-0.21
13	-0.02
15	0.33
12	0.003

TABLE 20 – Correlation BTC / Community

We can see that the Community 1 has the higher correlation coefficient with the price of Bitcoin, with the Louvain method, the Community wich **Poloniex.com** was part of had a correlation of 0.71, but with the Girvan-Newman method, the correlation coefficient is 0.77. So we might think that the Girvan-Newman method regrouped the actors in a more efficient way than the Louvain method.
However there is no significant difference between the correlation coefficient of the Community 1. Unfortunately, most of the communities computed with the Girvan-Newman method are lowly correlated with the price of Bitcoin.

We said in section *Community Evolution : Girvan_Newman Method*, that the communities computed were much more correlated with each other than the communities computed with the Louvain method.

If we look at the correlation matrix [22], the Community 1 was correlated with the Community 12 with a coefficient of 0.79. However, the Community 12 has a correlation coefficient of 0.003 with the price of Bitcoin. The problem is that the Community 12 started its activity in 2015-11 and the data of the price of Bitcoin that we have end in 2015-12.

We have this problem with Community 10, 7, 11, 13, 14, 12, 17, 9, 16, 15. So we can't really say that the communities computed with the Girvan-Newman method are correlated with the price of Bitcoin. However we are sure that monitoring the activity of the Community 1 and the key actor **Poloniex.com** could help us predict the trend of the price of Bitcoin be it with the Louvain method or the Girvan-Newman method.

5 Annexes

5.1 Community Detection

5.1.1 Community Detection : Louvain Method

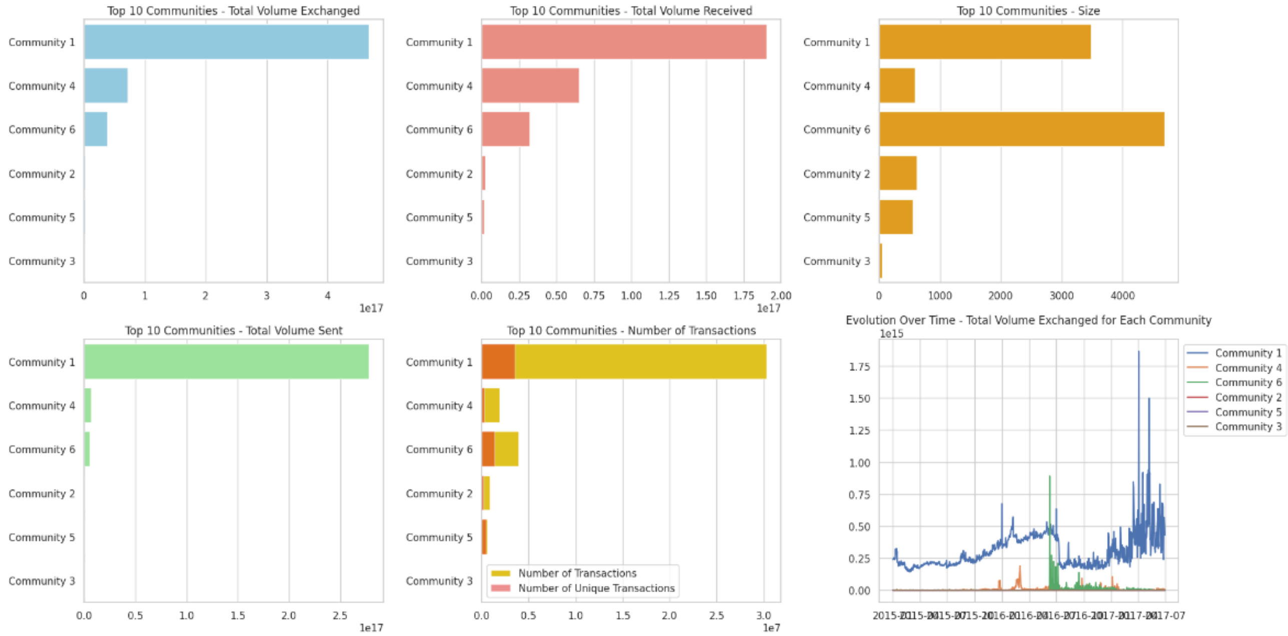

FIGURE 27 – Community Detection Louvain : Overview of the Network

5.1.2 Community Detection : Girvan-Newman Method

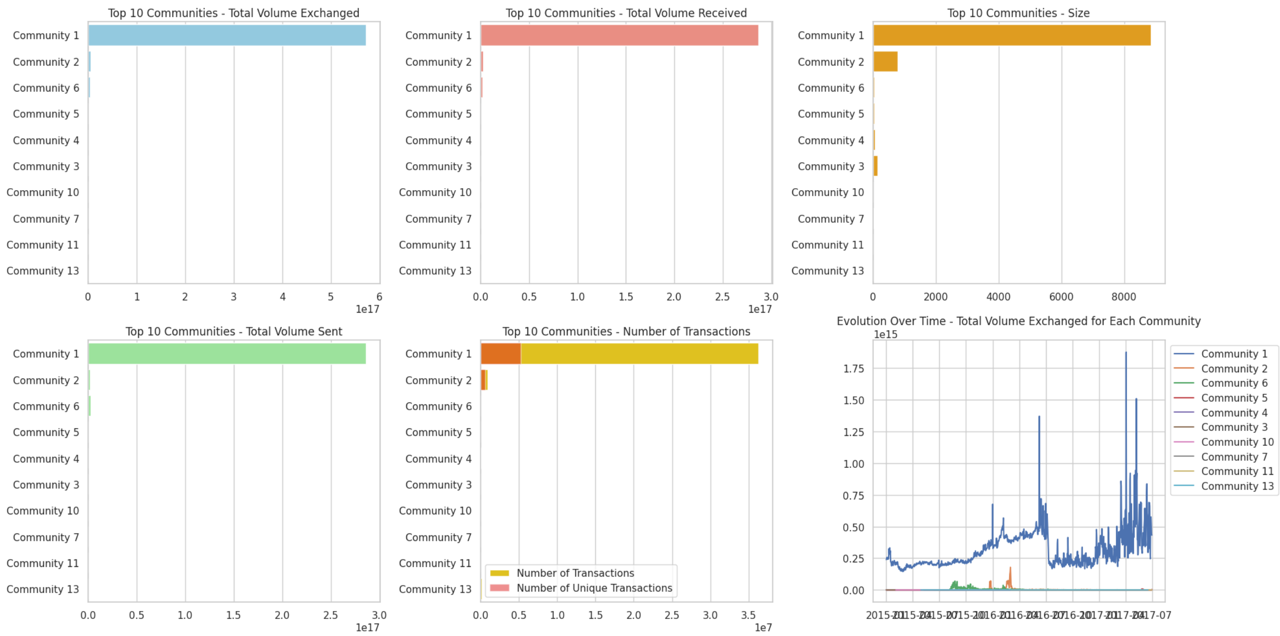

FIGURE 28 – Enter Caption

5.2 Communities Key Actors

5.2.1 Communities Key Actors : Louvain Method

FIGURE 29 – Communities Key Actors : Overview of the Network

5.2.2 Communities Key Actors : Girvan-Newman Method

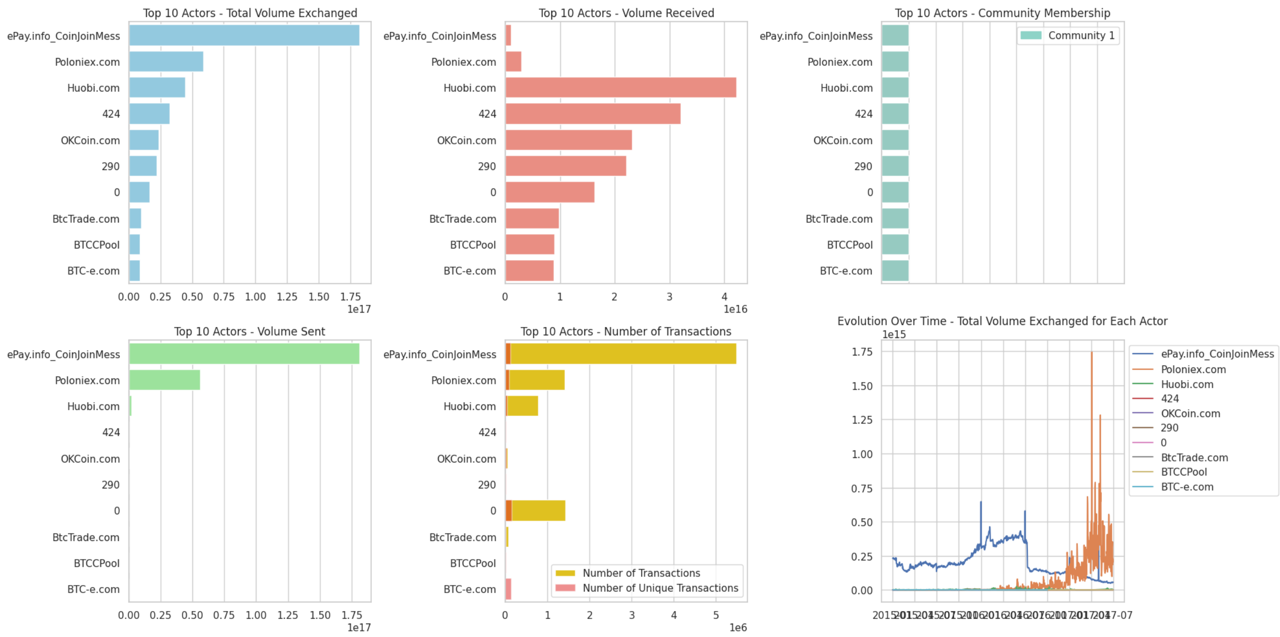

FIGURE 30 – Communities Key Actors : Overview of the Network

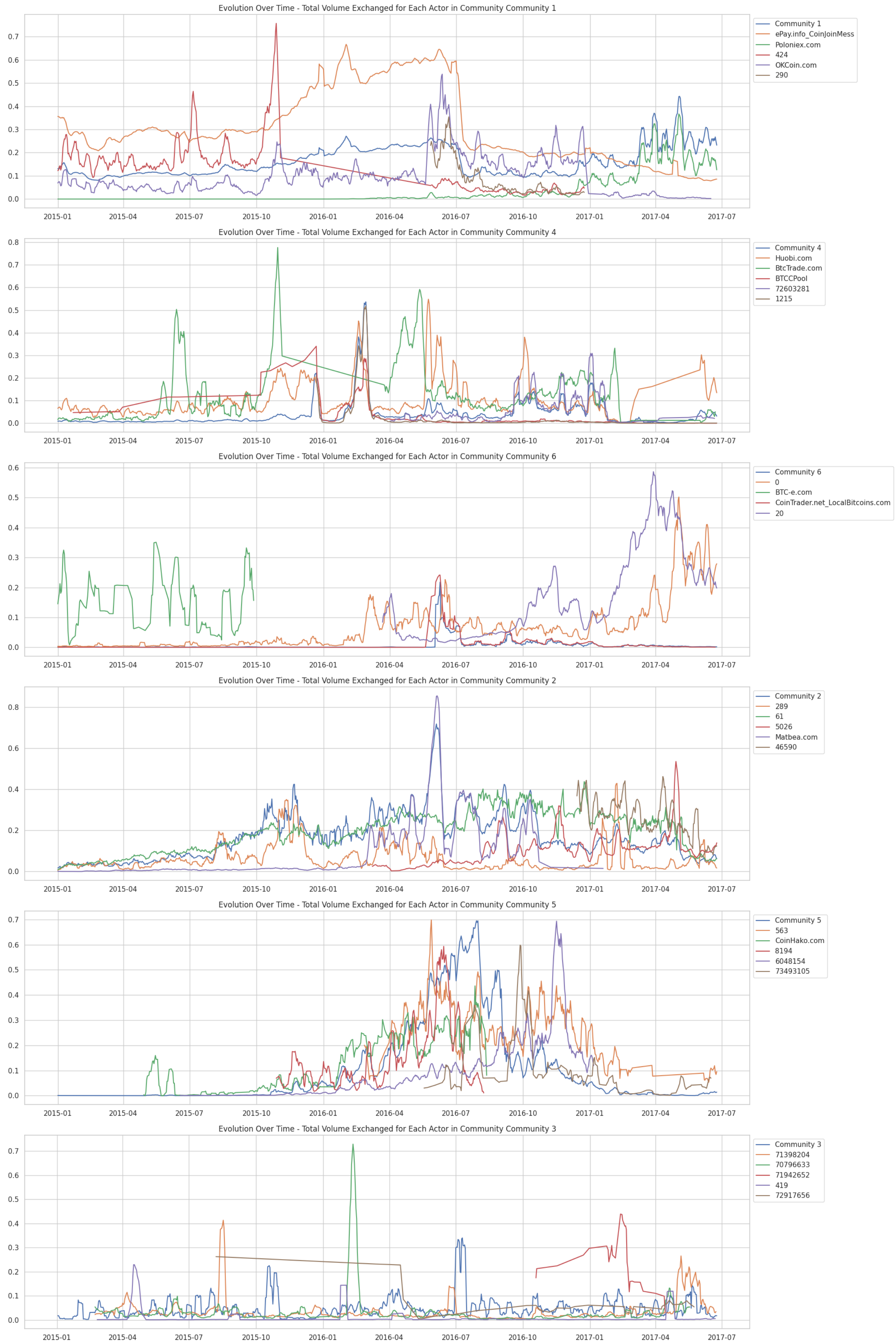

FIGURE 31 – Comparison of the Transaction Volumes of the Top Ten Actors for each Community

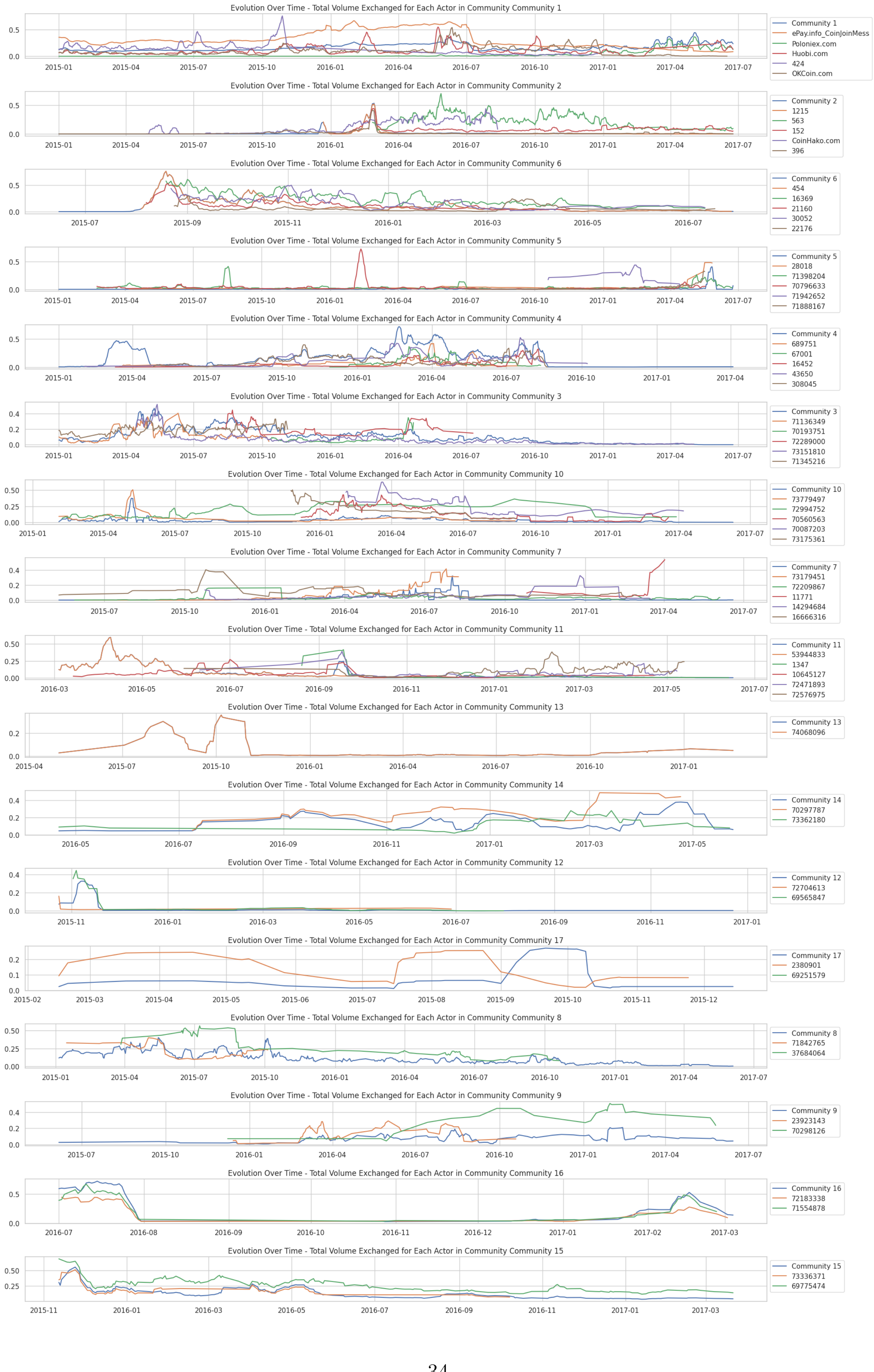

Figure 32 – Comparison of the Transaction Volumes of the Top Ten Actors for each Community

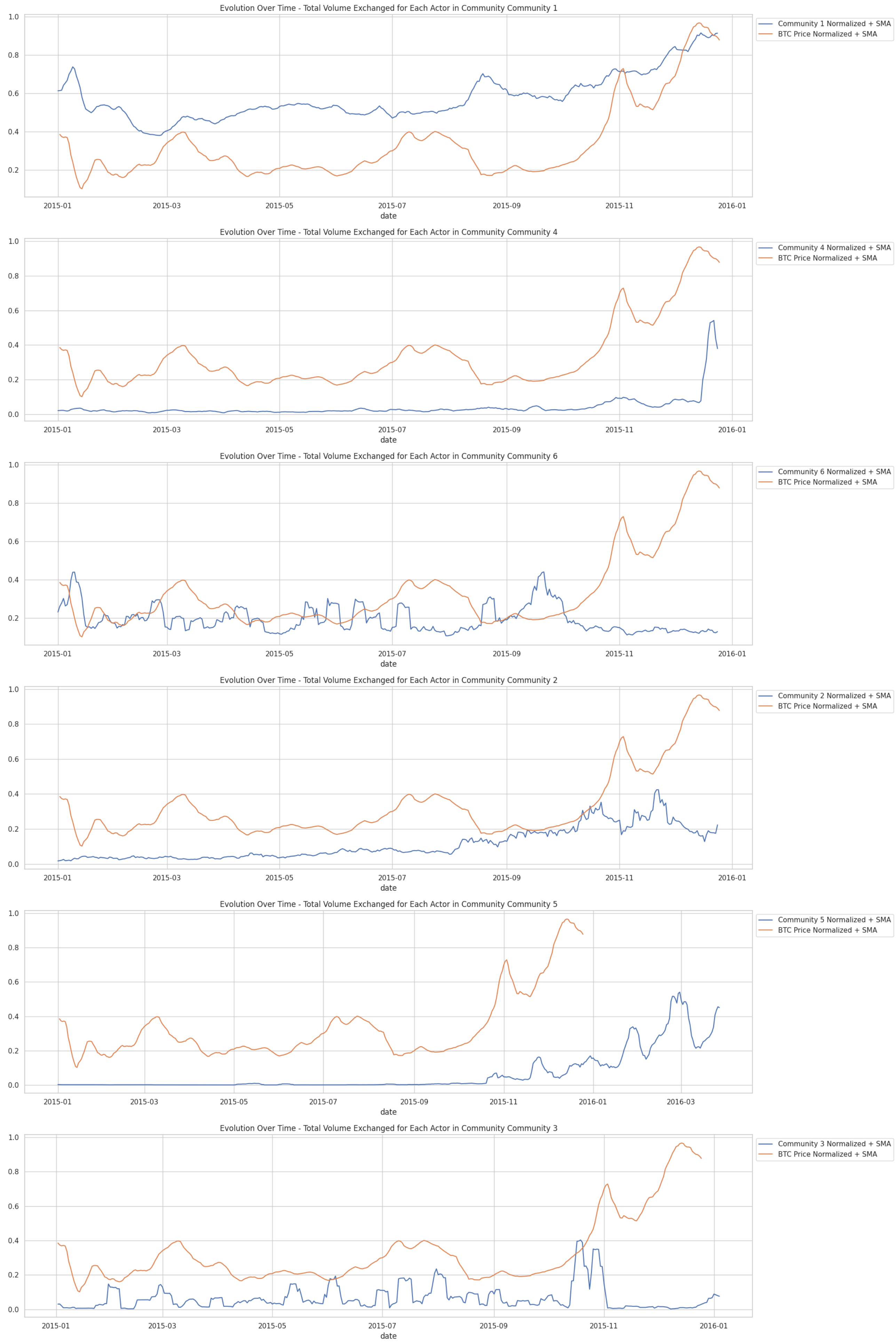

FIGURE 33 – Comparison of the Bitcoin Price and the Transaction Volumes of the Communities

FIGURE 34 – Evolution over Time - Total Volume Exchanged for Each Actor In Community X