Cryptocurrency Portfolio Diversification Using Network Community Detection

Dimitar Kitanovski, Miroslav Mirchev, Ivan Chorbev, and Igor Mishkovski
Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University in Skopje, North Macedonia
{dimitar.kitanovski, miroslav.mirchev, ivan.chorbev, igor.mishkovski}@finki.ukim.mk

***Abstract*—As of the end of 2013 till now we are witnessing huge volatility and risk in the cryptocurrency market compared to flat currency or stock market. Thus, in this market the portfolio diversification is of big importance in order to reduce volatility and keep the optimal return for the investors. A usual approach for portfolio construction is to keep a balance between returns and volatility, based on their interdependence and individual returns. One way of diversification is employing clustering or community detection algorithms to select a more diverse set of assets. We study the utilization of the Louvain algorithm and affinity propagation for community detection, based on correlation and mutual information between cryptocurrencies, for potential application in portfolio diversification.**

***Keywords*—Cryptocurrencies, Portfolio selection, Community detection, Financial analysis**

I. Introduction

Cryptocurrencies are decentralized electronic assets and payment systems that first surfaced in 2009, and have gained popularity as an investment alternative during the past several years [1]. Proportionately, instead of investing in conventional stocks and shares, substantial quantities of money have started to be placed in the cryptocurrency market over time. The cryptocurrency market, however, can see significant fluctuations due to a variety of factors, such as supply and demand, investor and user sentiment, regulatory limitations, and media excitement. Therefore, it is important to base the investment decisions in portfolio management on relevant analyses.

The study of cryptocurrency markets have many similarities with the study of other financial markets, which have been widely studied in the literature. The foundations of modern portfolio theory were set in [2], where the mean-variance framework was introduced, which balances between portfolios expected return and acceptable risk by imposing diversification. Financial markets can be represented as networks by calculating distances based on their correlations, as described in [3], where a minimum spanning tree (MST) representation was first introduced. Since then, there is an abundance of works in the areas of correlations, clustering, hierarchies and networks in financial markets, as summarized in [4].

Several recent works have studied correlations, network representations and community detection in cryptocurrency markets [5]–[8], including portfolio diversification. However, these studies typically apply correlation, which can have limitations as it only represents linear dependencies. Furthermore, most works use static community structures although in the dynamical crypto market correlations can change quickly. Therefore, in this study we employ community detection of a more dynamical network representation and test if mutual information can be used for capturing the relationships between cryptocurrencies, as it also depicts non-linear dependencies.

In Section II we briefly describe our dataset. Section III provides some initial relationship analyses in cryptocurrency markets, while in Section IV we describe how cryptocurrency networks can be build. Community detection in such networks is given in Section V, while in Section VI we examine how previous analyses can be applied to portfolio diversification.

II. Dataset

We collected historical cryptocurrency coins market price data from www.coinmarketcap.com using a Python scraper called *cryptocmd* for the period from 1/5/2013 to 26/3/2022. The dataset provides a wide variety of opportunities for additional analysis and investigation, as it includes two crucial historical periods that had an impact on the whole world and the global economy. The first one corresponds to the Covid-19 pandemic, which brought instabilities to all industries and economies worldwide. The second time period is the beginning of the global crisis brought on by the conflict between Russia and Ukraine, but we have not included it in this study.

This analysis included data from three consecutive years (2019, 2020, 2021), as well as cryptocurrencies with daily price records across the three years. The Pandas data analysis Python package was used to extract the daily closing prices (the price at the end of the day) for each cryptocurrency. The data required a small amount of data cleaning. Some irrelevant "disruptive" cryptocurrencies had extreme fluctuations in a

The authors would like to thank the Faculty of Computer Science and Engineering in Skopje for the financial support under project EPMAI.

very short time, so they were discarded from the dataset by filtering the cryptocurrencies by a market capitalization threshold of $2M. Cryptocurrencies that did not have price records for each day of the 3-year period were discarded, resulting in a total of 215 coins out of a total of 1000 coins.

III. Cryptocurrencies relationship analysis

Traditionally relationships between assets are analyzed using their cross-correlation, which can be used to build a correlation network [3]. However, correlation can only grasp their linear dependency, and some studies have shown that non-linear measures such as mutual information can be more informative [9]. Therefore, we conducted analysis of the cryptocurrencies relationships using both correlation and mutual information.

Instead of working directly with the daily price data series, they are first pre-processed by calculating the normalized logarithmic returns, which are more appropriate for analysis of financial assets due to their growth component. Let us denote the price of the i$^{\text{th}}$ cryptocurrency at time t by $p_i(t)$. The logarithmic returns $l_i(t)$ are calculated by taking the natural log of the difference between the prices of two consecutive days, i.e. $l_i(t) = p_i(t) - p_i(t - \Delta t)$, where Δt is a time period which in our case is 1 day. In order to standardize the volatility amongst cryptocurrencies, we normalize the logarithmic return values $l_i(t)$ by subtracting its mean $\overline{l_i}$ and dividing by its standard deviation σ_{l_i}, i.e. the normalized log return is $n_i(t) = (l_i(t) - \overline{l_i})/\sigma_{l_i}$. A detailed description of the data pre-processing process is available in [8].

A. Correlation Analysis

Finding linear relationships between cryptocurrencies is a major goal of the correlation matrix analysis. Pearson's correlation coefficient is used to calculate the correlation matrix $C_{N\times N}$, where $N = 215$ is the number of cryptocurrencies and $C_{ij} = \frac{\sum_t (n_i(t)-\overline{n_i})(n_j(t)-\overline{n_j})}{\sqrt{\sum_t (n_i(t)-\overline{n_i})^2 \sum_t (n_j(t)-\overline{n_j})^2}}$. The values of the correlation matrix vary from -1 to 1, where $C_{ij} = 1$ indicates a perfectly positive correlation between the cryptocurrencies, C_{ij} = -1 a completely negative correlation, and $C_{ij} = 0$ shows no correlation. The distribution of the coefficients in the cryptocurrency correlation matrix is given in Figure 1.

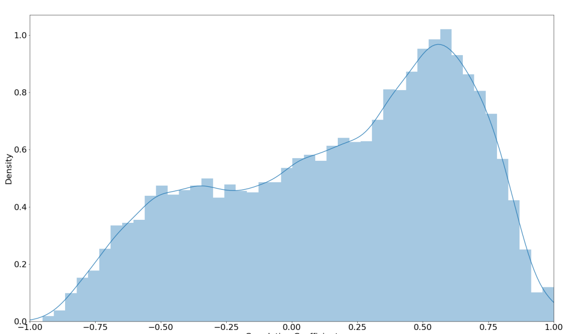


Fig. 1. Distribution of coefficients in the cryptocurrency correlation matrix.

A technique called Random matrix theory (RMT) can be applied on the obtained correlation matrix in order to validate that it carries relevant information, as it was also demonstrated in [8]. It involves a comparison of the distribution of eigenvalues of the correlation matrix C with that of a random matrix R. Our analysis of a random matrix R with the same dimension have shown analytically that its eigenvalues fall within the range $\lambda_R = (\lambda_-, \lambda_+)$, where $\lambda_- = 0.3104$, and $\lambda_+ = 2.1526$ are the expected lower and upper bounds of the eigenvalues. In Figure 2, it can be seen that many eigenvalues of the correlation matrix C are outside the range λ_R, denoted with vertical blue lines. Therefore, we can state that the correlation matrix captures useful information about the relationships between the cryptocurrencies.

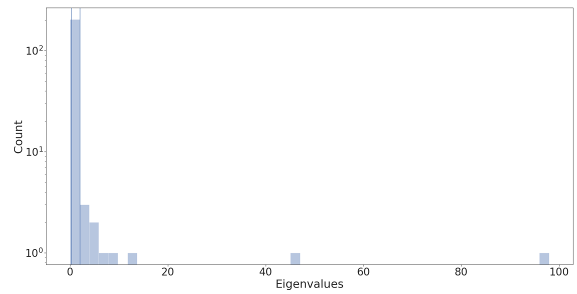


Fig. 2. Eigenvalues distribution of the cryptocurrency correlation matrix.

B. Mutual Information Analysis

In addition to the linear dependencies obtained from the correlation matrix, the non-linear dependencies between the cryptocurrencies were also extracted by calculating a mutual information matrix M between cryptocurrencies to get a more precise overview of the cryptocurrencies dependencies. The linear dependencies can be insufficient to completely describe the relationships between cryptocurrencies over time since the market is a very volatile environment. A large number of cryptocurrencies are not correlated at all according to the the correlation matrix, while the same cryptocurrencies are highly related according to the mutual information matrix. The pairs of cryptocurrencies like (BTCBitcoin, HNDCHondaisCoin) and (BLAZRBlazerCoin, HNDCHondaisCoin) are some of the cryptocurrencies that belong to a category of currencies that have no relationship according to the calculated linear dependencies yet have a strong relationship according to the information from the calculated nonlinear dependencies.

The mutual information (MI) of two cryptocurrencies is expressed as $M_{ij} = H(i) - H(i|j)$, where $H(i)$ stands for the price entropy of cryptocurrency i, $H(i|j)$ stands for its conditional entropy given cryptocurrency j, and $M_{ij} \in (0, \infty)$. In other words it tells how much more information we can get about i if we have information about j. Small value of M_{ij} means that there is a weak relationship between the cryptocurrencies, while large values indicate that they are highly related and there is a significant reduction in uncertainty. Finally, if M_{ij} = 0 it means that cryptocurrencies are completely independent of each other. Statistical analyses of the distribution of the values in the correlation and mutual information distance matrices were carried out to identify the differences between them, as shown in Figure 3.

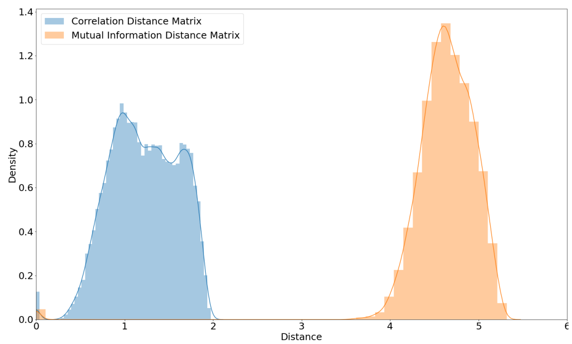


Fig. 3. Distribution of values in the correlation and MI distance matrices.

IV. Cryptocurrency networks

There are certain criteria that matrix values must fulfill to be used for network analysis and creation of a network structure. In [3], [8] the authors provide a thorough explanation of why directly using the correlation matrix is inappropriate for creating a network representation. Instead a distance metric is typically employed to derive a corresponding distance matrix. Different procedures are required for producing distance matrices from the correlation and mutual information matrices.

A positive distance matrix D can be created from the correlation matrix C using a transformation [3], [5], [7], [8]

$$D_{ij}^{C} = \sqrt{2(1 - C_{ij})}, \tag{1}$$

where $D_{ij}^{C} = 0$ indicates perfectly correlated coins i and j, $D_{ij}^{C} = 2$ shows completely negatively correlated coins, while $D_{ij}^{C} = 1.41$, signifies no correlation between the coins.

To obtain a distance matrix D from the mutual information matrix M we can use a simple transformation

$$D_{ij}^{M} = |M_{ij} - \max(M_{ij})|. \tag{2}$$

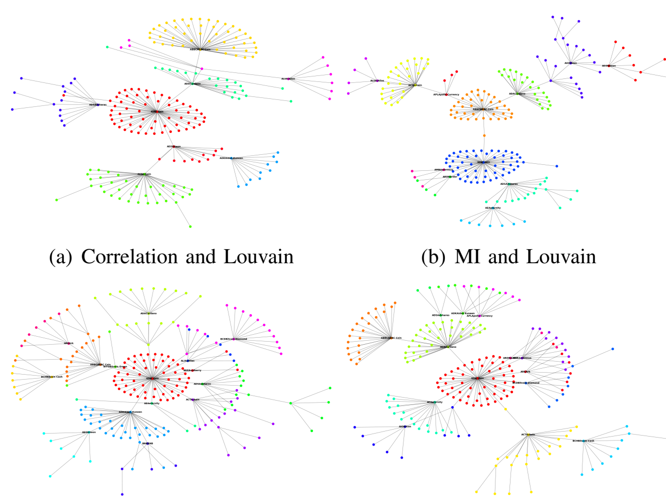
The meaning of the values of the transformed matrix are changed, hence, large values of D_{ij}^{M} indicate that there is weak relation between the coins i and j, while small D_{ij} means that the coins i and j are strongly related.

V. Community detection

In this work we use the affinity propagation (AP) and Louvain community detection (clustering) algorithms. AP's main advantage is its ability to accept negative values (weights), thus it can be applied directly on the correlation and MI matrices. AP is an iterative clustering algorithm where information is propagated between data points in space, i.e. nodes in our case, in two phases of responsibility and availability for a predefined number of iterations or until no change is detected. The Louvain community detection algorithm, on the other hand, is a network-based algorithm that forms communities using a "bottom-up" approach. The network's modularity is the primary indicator employed in the background of combining data points. The Louvain method's primary goal is to establish communities that will be distinguished by a high level of modularity. With both AP and Louvain the number of communities does not need to be predetermined.

An investigation of the stability of the communities was carried out on a monthly basis throughout the course of the 3-year period. An overlapping coefficient between the communities was calculated from the monthly data and its value varies from 20% to 35%. Such a conclusion demonstrates the unpredictability of the cryptocurrency market. The foregoing approaches produce very volatile communities, and the relationships between the cryptocurrencies themselves alter significantly on a monthly basis. Therefore, a co-occurrence matrix was built with elements representing how often two cryptocurrencies belong to the same community, which was then used to build a cryptocurrencies network, in order to amortize the high instability within communities. The co-occurrence matrix is an $N \times N$ matrix, where N is the total number of cryptocurrencies (in this study $N = 215$).

In the next section we examine communities both with and without a co-occurrence transformation, but here we visualize only communities obtained from the co-occurrence matrices. We employ two distinct strategies to apply the community detection methods. AP is applied directly to the co-occurrence data obtained from the correlation and MI matrices, while Louvain was carried out using the Minimum Spanning Tree (MST) produced from the co-occurrence weighted network. The relationships between cryptocurrencies are represented as a weighted network, where the weights are derived from the co-occurrence matrices which are calculated from the correlation and MI matrices. Figure 4 shows the minimum spanning tree (MST) of the cryptocurrency data derived using the aforementioned co-occurrence matrices, including the communities and leading coins that are obtained. The communities with AP are calculated on a fully connected weighted network, but the visualization is shown as a MST.



(a) Correlation and Louvain (b) MI and Louvain

(c) Correlation and AP (d) MI and AP

Fig. 4. Detected communities in correlation and mutual information (MI) networks with Louvain and Affinity propagation (AP) methods.

VI. Portfolio Diversification

After analyzing the networks produced by the correlation matrix (C), mutual information matrix (MI), and the derived co-occurrence matrices (C', MI'), the final step in this study is to apply the techniques for portfolio selection. The community detection methods are used for diversifying the portfolio by selecting cryptocurrencies from different communities. Two distinct approaches were used for the selection process, and for each approach several strategies with different "centrality" scores were examined. In the first approach principal component analysis (PCA) is used for an analysis of the communities of cryptocurrencies in order to characterize the variability of each cryptocurrency. The second approach, centrality degree (C_D), is a node level measure calculated as the quotient of the number of nearby nodes and the total number of network nodes. We then use several strategies to select the nodes with the PCA and C_D scores, namely maximum, median, and minimum, to produce a thorough examination. For the analyses in this section we used the same data from 2019 to 2021, but we split it into 12 one-year train and one-year test windows, by sliding through time by one month. The reported values are the median of these 12 repeated trials. Table I gives the annual returns (%) , and we can notice that the co-occurrence matrix can be beneficial.

TABLE I
PORTFOLIO ANNUAL RETURNS (%) FOR DIFFERENT STRATEGIES

		Louvain				**Affinity Propagation**			
		C	*MI*	*C'*	*MI'*	*C*	*MI*	*C'*	*MI'*
PCA	*max*	356	271	421	767	264	343	363	215
	med	253	346	357	329	315	264	405	245
	min	308	294	518	274	464	284	242	264
C_D	*max*	171	342	451	510	257	515	298	370
	med	328	357	339	358	320	268	416	351
	min	287	357	339	358	339	442	416	351

Table II shows the standard deviation of the portfolios log-returns over time in order to quantify the volatility to which the particular approach is susceptible, while in Table III the ratio between the returns in percents and the volatility is given.

TABLE II
PORTFOLIO VOLATILITY (STANDARD DEVIATION OF LOG-RETURNS)

		Louvain				**Affinity Propagation**			
		C	*MI*	*C'*	*MI'*	*C*	*MI*	*C'*	*MI'*
PCA	*max*	.66	.48	.56	.71	.52	.66	.58	.53
	med	.61	.58	.50	.55	.61	.56	.70	.55
	min	.55	.56	.65	.48	.58	.62	.51	.51
C_D	*max*	.53	.62	.59	.65	.62	.71	.45	.57
	med	.61	.50	.59	.58	.61	.62	.59	.58
	min	.47	.50	.59	.58	.60	.62	.59	.58

TABLE III
PORTFOLIO ANNUAL RETURNS DIVIDED BY ITS VOLATILITY

		Louvain				**Affinity Propagation**			
		C	*MI*	*C'*	*MI'*	*C*	*MI*	*C'*	*MI'*
PCA	*max*	539	565	752	1080	508	520	626	406
	med	415	597	714	598	516	471	579	445
	min	560	525	797	571	800	458	475	518
C_D	*max*	323	552	764	785	415	725	662	649
	med	538	714	575	617	525	432	705	605
	min	611	714	575	617	565	713	705	605

VII. Conclusion

In this paper we studied the application of two community detection (clustering) algorithms, Louvain and Affinity propagation, for portfolio construction in the cryptocurrency market. Moreover, we examined cryptocurrency relationships representation by correlation and mutual information matrices and distance matrices, as well as how we can benefit from the employment of a co-occurrence matrix which captures the market dynamics more closely.

From these preliminary results, we can not clearly conclude what is the best investment strategy. However, we can see that using the co-occurrence matrices can generally bring improvements, while using mutual information can give better results for some strategies, but worse for others. We continue our study with other network structures, community detection and selection methods, including graph neural networks as described in [10]. Furthermore, we plan to examine the application of different strategies for various time periods and market conditions.

References

[1] Laura Alessandretti, Abeer ElBahrawy, Luca Maria Aiello, and Andrea Baronchelli. Anticipating cryptocurrency prices using machine learning. *Complexity*, 2018.

[2] Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[3] Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.

[4] Gautier Marti, Frank Nielsen, Mikołaj Bińkowski, and Philippe Donnat. *A Review of Two Decades of Correlations, Hierarchies, Networks and Clustering in Financial Markets*, pages 245–274. Springer, Cham, 2021.

[5] Darko Stosic, Dusan Stosic, Teresa B Ludermir, and Tatijana Stosic. Collective behavior of cryptocurrency price changes. *Physica A: Statistical Mechanics and its Applications*, 507:499–509, 2018.

[6] Theophilos Papadimitriou, Periklis Gogas, and Fotios Gkatzoglou. The evolution of the cryptocurrencies market: A complex networks approach. *Journal of Computational and Applied Mathematics*, 376:112831, 2020.

[7] Harshal Chaudhari and Martin Crane. Cross-correlation dynamics and community structures of cryptocurrencies. *Journal of Computational Science*, 44:101130, 2020.

[8] Jenna Gavin and Martin Crane. Community detection in cryptocurrencies with potential applications to portfolio diversification. *arXiv preprint arXiv:2108.09763*, 2021.

[9] Alexander Haluszczynski, Ingo Laut, Heike Modest, and Christoph Räth. Linear and nonlinear market correlations: Characterizing financial crises and portfolio optimization. *Physical Review E*, 96(6):062315, 2017.

[10] Martin JA Schuetz, J Kyle Brubaker, and Helmut G Katzgraber. Combinatorial optimization with physics-inspired graph neural networks. *Nature Machine Intelligence*, 4(4):367–377, 2022.