Interpretable Machine Learning using Visibility Graph and Random Forests

Andrii Bielinskyi[1,2,*,†], Vladimir Soloviev[1,3†], Andriy Matviychuk[1,†] and Halyna Velykoivanenko[1,†]

[1] *Kyiv National Economic University named after Vadym Hetman, 54/1 Beresteysky Ave., Kyiv, 03057, Ukraine*
[2] *State University of Economics and Technology, 16 Medychna Str., Kryvyi Rih, 50005, Ukraine*
[3] *Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine*

Abstract

We examine whether network representations of financial series produce interpretable predictive signals. Daily S&P 500 prices are mapped to Natural Visibility Graph (NVG), from which we extract multi-scale topological and spectral descriptors using overlapping windows of 100, 250, and 500 trading days. These features drive Random Forest (RF) models for two 7-day-ahead tasks: (i) directional classification (up/down) and (ii) magnitude regression of standardized forward returns, with training and evaluation conducted in temporal order. RFs are used for their robustness to heterogeneous inputs and their built-in mean decrease in impurity (MDI), enabling direct ranking of NVG features by contribution to performance. Out-of-sample, the classifier attains ROC–AUC $= 0.62$ and accuracy ≈ 0.584 on balanced classes – statistically meaningful yet economically modest. Beyond point accuracy, the approach yields transparent importance profiles that identify which NVG attributes are most informative for short-horizon forecasts. Overall, the evidence indicates that VG features provide complementary, structure-aware information for stock-index prediction while preserving interpretability through RF-based importance analysis.

Keywords

Visibility graph, complex networks, Random Forest, S&P 500, feature importance, interpretable machine learning, network measures, sliding window

1. Introduction

Financial indexes like the S&P 500 display rich, nonlinear behavior that challenges conventional forecasting. A growing line of work uses complex network theory to represent time series as graphs, enabling structural analysis across scales [1]. In particular, the visibility graph (VG) maps each time point to a node and links pairs that have a direct line of sight in the time–value plane, translating geometric relations into network topology [2]. This representation preserves key properties of the signal and supports high-level descriptors – e.g., connectivity and clustering – that summarize temporal dynamics [3, 4]. In short, network mappings let us study financial series with graph-theoretic tools, yielding features that capture temporal complexity beyond standard statistics.

The rationale for network-based forecasting is to expose structural information that time-domain methods can miss. Measures such as average degree, clustering, spectral radius, and path-length statistics quantify the topology of a series' VG and relate to phenomena like volatility clustering and

Information Technology and Implementation (IT&I-2025), November 20-21, 2025, Kyiv, Ukraine
* Corresponding author.
† These authors contributed equally.
bielinskyi@kneu.dp.ua (A. Bielinskyi); vnsoliviev@kdpu.edu.ua (V. Soloviev); matviychuk@kneu.edu.ua (A. Matviychuk); ivanenko@kneu.edu.ua (H. Velykoivanenko)
0000-0002-2821-2895 (A. Bielinskyi); 0000-0002-4945-202X (V. Soloviev); 0000-0002-8911-5677 (A. Matviychuk); 0000-0001-6326-3965 (H. Velykoivanenko)

cyclicity [1, 3]. Prior studies show that VGs retain essential dynamics and produce discriminative features for prediction [3, 5]; for equity indices, VG-based analysis has delivered signals at both short and longer horizons [5, 6]. This aligns with a broader trend of combining topological data analysis with machine learning (ML), since network-derived features complement traditional predictors and enrich the feature set [1, 9].

At the same time, interpretability has become central in ML – especially in finance, where understanding model rationale is critical. Black-box models (e.g., deep networks) can obscure decision drivers and raise risk concerns [7]. Random Forests (RFs) offer a middle ground: competitive accuracy together with built-in feature importance that quantifies each predictor's contribution to error reduction across trees [8]. Such transparency is valuable for analysts and regulators who need to know which inputs drive forecasts. RFs therefore bridge complex, data-driven analysis and the need for explanation.

Motivated by these developments, we study RFs as interpretable rankers of VG-derived features for financial forecasting. Using the S&P 500, we construct VGs on sliding windows of 100, 250, and 500 trading days to capture evolving structure [2, 3]. From each window we extract topological descriptors (average degree, clustering, spectral radius, path-length metrics), then train RFs for two tasks: (i) regression of the 7-day forward return and (ii) classification of 7-day direction. Our objective is not to surpass forecasting benchmarks, but to use RFs to identify which VG features carry the strongest predictive signal.

In summary, we integrate complex network representations with interpretable ML to advance financial time-series modeling. Representing an index as a VG yields a spectrum of structural features [2–4]; RF analysis then highlights the most informative among them for short-term forecasting [8]. The approach ties time-series network characteristics to later market movements and provides a practical, feature-oriented guide for prioritizing the most impactful complex-network features.

2. Literature Review

2.1. Network-Based Approaches in Financial Market Analysis and Forecasting

VGs map a time series into a network by linking samples that satisfy a line-of-sight criterion, preserving salient geometric dynamics of the original signal [1, 2, 10]. For financial data, VGs uncover scale-free degree distributions and long-range dependence (e.g., global indices with power-law scaling), and VG-based metrics – degree statistics, entropy, clustering – capture nonlinear patterns that traditional statistics may miss; they have also been used for characterization and prediction of price movements [11–13].

Complementary to VGs, correlation-based market networks (and their minimum spanning trees, MSTs) reveal hierarchical structures (e.g., sectoral clustering) and trace regime shifts: during crises networks densify and lose modularity; in tranquil phases they are sparser and more fragmented [14–18]. Such topology shifts support regime detection and systemic-risk analysis, while spillover and co-movement graphs help model contagion across markets [19, 20].

Beyond description, forecasting the network itself (e.g., correlation-link addition/removal) via ML with node/edge features improves predictive accuracy over raw correlations and enhances portfolio rebalancing and risk control [21, 22]. Network centrality also informs allocation: favoring peripheral (low-centrality) assets tends to improve diversification and stabilize performance in stress periods [23].

Recent work leverages graph neural networks (GNNs) to jointly learn temporal and cross-sectional dependencies. Spatio-temporal GAT variants (e.g., FSTGAT) have outperformed LSTM/XGBoost baselines, particularly in volatile regimes, and can anticipate turning points by learning dynamic inter-asset relations – though interpretability remains a key concern prompting research on explainable GNNs [24–29].

Recent work integrates VG features ML models for time-series forecasting and classification, leveraging network metrics – such as degree, path length, and centrality – as informative representations of temporal patterns [30, 31]. In finance, studies like these [32, 33] showed that VG-derived topological indicators rise during market turbulence, enabling classifiers such as SVM and k-NN to predict next-day volatility with over 70% accuracy. Yao similarly used VG metrics with logistic regression to classify stock valuations, and Kutluana et al. applied weighted VGs to heartbeat data for medical prediction tasks [34, 35]. These examples underscore VG features' utility across domains, particularly in interpretable ML settings [36, 37].

Among ML approaches, RFs are especially well-suited for VG-based forecasting due to their flexibility and built-in feature importance metrics. Studies by Singh et al. and Bocaccio et al. have demonstrated improved accuracy using RFs with VG inputs in contexts ranging from financial time series to audio signal classification [38, 39]. In finance, RFs not only match or exceed traditional models in forecasting tasks (e.g. S&P 500 direction) but also clarify which VG metrics – like clustering or degree heterogeneity – most influence predictions [40, 41].

Network-based methods – from correlation graphs to GNNs – have proven effective in capturing complex dependencies in financial data that traditional models often miss [26, 28, 30, 42]. This study contributes to that paradigm by using VGs to transform individual stock time series into networks, enabling the extraction of topological features (e.g., degree, clustering, motifs) that reveal volatility and structural complexity [1–3, 33, 37, 41]. VG features thus offer intra-series analogs to inter-asset correlation networks, encoding rich temporal dynamics.

We extend this framework by pairing VG-derived features with RF, emphasizing interpretability alongside predictive performance. Unlike black-box models such as GNNs, RFs provide transparent insights via feature importance scores, supporting explainable AI in finance. Our approach bridges descriptive VG analysis and interpretable forecasting, showing how structural features of price series can improve predictions while revealing which patterns matter most. By doing so, we advance structure-aware modeling with a method that is both rigorous and accessible for practical decision-making.

3. Research Methodology

3.1. Visibility Graph Representation

We generate VGs from time series data using the natural VG (NVG) algorithm, following the approach introduced in the original study [2]. In this method, each time point i becomes a node, and two nodes $i < j$ are connected if there is an unobstructed "line of sight" between them – i.e., if every intermediate point k with $i < k < j$ satisfies:

$$x_k < x_j + \left(x_i - x_j\right)\frac{j-i}{j-k}.$$

This condition ensures the straight line connecting points (i, x_i) and (j, x_j) lies above all intermediate points (k, x_k), establishing visibility. Applying this rule, we construct an undirected VG $G = (V, E)$, where $|V| = N$ nodes correspond to the time series length. Due to the nature of the construction, each VG is fully connected across consecutive time steps. The resulting binary, symmetric adjacency matrix A encodes the structural profile of the time series as a complex network. All further network-based analysis is conducted on these graphs.

3.2. Spectral and Topological Measures of Network Structure

Once the VGs are constructed, we compute a suite of topological and spectral metrics to quantify their structural properties.

Clustering Metrics: *Global Clustering Coefficient (C)* captures the likelihood of triangle formation by averaging local clustering over all nodes: $C = N^{-1}\sum_{i=1}^{N} 2E_i/k_i(k_i - 1)$ where E_i is the

number of connections among node i's neighbors. *Transitivity (T)* measures the ratio of closed triplets to all connected triplets, indicating overall triangle density. *Square Clustering* evaluates the frequency of 4-node cycles, reflecting square-like substructures in the graph.

Efficiency and Path Length: *Global Efficiency* E_{glob} is the average inverse shortest-path length across all node pairs, indicating network-wide navigability. *Local Efficiency* E_{loc} captures the efficiency within each node's neighborhood subgraph, assessing fault tolerance. Small-world networks typically exhibit high E_{loc} and low average path length, combining local clustering with global connectivity.

Assortativity (r): Measures the correlation of node attributes (e.g., degree) at both ends of an edge. Positive r implies similar nodes connect; negative r implies dissimilarity.

Centrality and Hubs: *Maximum Degree* k_{max} indicates the most connected node, suggesting dominant time points. *Betweenness Centrality* b_k quantifies how often a node lies on shortest paths, highlighting influential "bridges" in the time series that may correspond to market regime shifts.

Small-Worldness (S): Defined as $(C/C_r)/(L/L_r)$, where C_r and L_r are clustering and path length in a random graph. $S > 1$ indicates small-world structure – high local clustering with short global paths.

Spectral Measures: *Graph Index Complexity (GIC)* uses the spectral radius λ_{max} of the adjacency matrix to capture structural complexity: $GIC = 1 - (2c - 1)^2$, where c is the normalized spectral radius. GIC peaks for intermediate connectivity, distinguishing graphs that are neither too sparse nor too dense. *Algebraic Connectivity* λ_2 (second-smallest Laplacian eigenvalue) reflects overall network cohesion and robustness. *Adjacency Spectral Gap* $\delta_A = \lambda_1 - \lambda_2$ assesses the dominance of the leading eigenmode; a large gap suggests integration, while a small gap indicates community structure.

Together, these metrics characterize both local motifs and global architecture of the VG, offering a multi-scale view of the time series' underlying structure.

3.3. Random Forest for Regression and Classification

RF are ensemble models that combine multiple decision trees to enhance prediction accuracy and reduce overfitting. In classification, each tree votes, and the majority class is chosen; in regression, predictions are averaged [8]. By aggregating trees trained on bootstrapped data and using random subsets of features at each split, RF reduces variance and improves generalization compared to single decision trees.

Each tree is built using recursive binary splits, selecting the feature and threshold that maximize impurity reduction (e.g., Gini index or variance). Splitting continues until a stopping criterion is met (e.g., max depth or minimum node size). This results in diverse trees that capture complex patterns while ensemble averaging controls overfitting.

Key RF hyperparameters include:

- **n_estimators**: Number of trees (e.g., 100–500). More trees generally reduce variance, with diminishing returns beyond a point.
- **max_features**: Number of features considered at each split. Smaller values increase diversity but may raise bias. Defaults are √p for classification and full feature sets for regression.
- **max_depth**: Maximum tree depth. Fully grown trees (no limit) are common, but limiting depth can reduce complexity and overfitting.
- **min_samples_split/leaf**: Minimum number of samples to split or form a leaf, used to regularize overly deep trees.
- **bootstrap**: Whether to use bootstrapped samples. When enabled (default), it increases diversity and allows for out-of-bag error estimation.

Additional parameters like max_leaf_nodes or n_jobs aid in controlling tree size or parallelizing training. In this study, RF is implemented using Scikit-learn, and hyperparameters are tuned empirically via validation or randomized search [43].

3.4. Tree's Feature Importance from Mean Decrease in Impurity (MDI)

RFs provide embedded feature importance via the MDI – often called Gini importance when using the Gini index [8]. For feature X_j, its importance is the average, over all M trees, of the impurity reductions from splits on X_j, weighted by the node sample fraction:

$$\mathrm{Imp}(X_j) = \frac{1}{M}\sum_{m=1}^{M}\sum_{\substack{t\in T_m:\\ v(s_t)=X_j}}\frac{N_t}{N}\Delta i(t)\,.$$

Importances are non-negative and typically normalized to sum to 1. MDI is fast and useful for ranking predictors and for feature selection; averaging across many randomized trees also stabilizes estimates [44].

MDI is computed on training splits and can overstate importance; validate selections with cross-validation. It also favors high-cardinality/continuous features [45]. As a check, use permutation importance (mean decrease in accuracy) on held-out data, which is more computationally costly but less biased [46]. In practice, combine MDI (quick heuristic) with permutation tests (robust verification) for reliable, interpretable feature ranking.

3.5. Time Series Preprocessing and Sliding Window Analysis

We address long-horizon nonstationarity with an overlapping sliding-window scheme. For daily S&P 500 prices S_t and window lengths $w \in \{100, 250, 500\}$, each segment $\mathcal{W}_{t,w} = \{S_{t-w+1}, \dots, S_t\}$ for $t = w, \dots, T$ is mapped via the natural-visibility rule to a VG $G_{t,w}$. From each $G_{t,w}$ we compute a descriptor vector $\varphi(G_{t,w}) \in \mathbb{R}^p$; concatenating scales yields a multi-scale feature stream

$$X_t = \left[\varphi(G_{t,100})^{\mathrm{T}}, \varphi(G_{t,250})^{\mathrm{T}}, \varphi(G_{t,500})^{\mathrm{T}}\right]^{\mathrm{T}}.$$

Overlapping windows smooth feature evolution and enable high-resolution tracking of structural change. Crucially, features at time t use only $\mathcal{W}_{t,w}$ (no look-ahead). The sample spans 23 Dec 1981–21 Aug 2025.

Targets use a 7-day horizon $h = 7$. The forward return is

$$R_{t,h} = (S_{t+h} - S_t)/S_t.$$

standardized with a trailing 50-day window:

$$\mu_t = \tfrac{1}{50}\textstyle\sum_{s=t-50}^{t-1} R_{s,h},\ \sigma_t = \sqrt{\tfrac{1}{49}\textstyle\sum_{s=t-50}^{t-1}\left(R_{s,h} - \mu_t\right)^2},\ r_{t,h} = \left(R_{t,h} - \mu_t\right)/\sigma_t.$$

We use $r_{t,h}$ for regression and $y_{t,h} = \mathrm{sign}(r_{t,h})$ for classification. Standardization mitigates drift/volatility shifts and helps balance Up/Down classes.

3.6. Hyperparameters Tuning

We tune hyperparameters $\theta \in \Theta$ via randomized search, which explores high-impact ranges efficiently without the combinatorial cost of grid search. Evaluation is time-aware: a 5-split purged, expanding cross-validation with an embargo $g = 500$ days (equal to the maximum feature window $w_{max} = 500$ prevents leakage from overlapping windows.

Let $1 < \tau_1 < \cdots < \tau_K < T$. For split k: train $= [1, \tau_k]$, embargo $= (\tau_k, \tau_k + g]$, test $= (\tau_k + g, \tau_{k+1}]$. For each candidate $\theta^{(b)}$ and split k, we fit on the train set and score on the test set using the task-appropriate metric. The cross-validated objective is

$$\hat{J}(\theta) = \frac{1}{K}\sum_{k=1}^{K} \text{Score}_k(\theta),$$

and we select $\theta^* = \text{argmax}_{\theta\in\Theta}\,\hat{J}(\theta)$ (or argmin, depending on the metric).

This procedure emulates real-time deployment and strictly enforces temporal separation. We run it independently per model class (e.g., RF with varied $n_{estimators}$, max_depth, max_features, etc.).

3.7. Scoring Metrics and Evaluation Performance

We use time-ordered samples $\{(x_t, r_{t,h})\}_{t=t_0}^{T-h}$, where $x_t \in \mathbb{R}^p$ are VG descriptors and $r_{t,h}$ is the h-day standardized return ($h = 7$). For direction, labels are $y_{t,h} = 1\{r_{t,h} > 0\} \in \{+1, -1\}$.

Modeling & tuning. Regression fits $r_{t,h} \approx f_\theta(x_t)$; we tune hyperparameters λ via purged, forward-chaining CV by minimizing the median absolute error (MedAE) across splits, then refit on the full training range. For classification, we maximize mean ROC–AUC across splits.

Error metrics (regression), with $e_i = y_i - \hat{y}_i$:

- $\text{MAE} = N^{-1}\sum_{i=1}^{N}|e_i|$;
- $\text{MSE} = N^{-1}\sum_{i=1}^{N} e_i^2$;
- $\text{RMSE} = \sqrt{\text{MSE}}$;
- $R^2 = 1 - \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{N}(y_i - \bar{y})^2$.

Classification metrics (TP, TN, FP, FN):

- $\text{ACC} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$;
- $\text{Prec} = \text{TP}/(\text{TP} + \text{FP})$;
- $\text{Rec} = \text{TP}/(\text{TP} + \text{FN})$;
- $\text{F1} = (2\ \text{Prec}\cdot\text{Rec})/(\text{Prec} + \text{Rec})$.

Report per-class scores and summarize by macro average $C^{-1}\sum_c \text{Metric}_c$ or weighted average $\sum_c (n_c/N)\text{Metric}_c$.

4. Empirical Results

This section evaluates the short-horizon predictive content of VG features for the S&P 500. We consider two 7-day tasks: (i) directional classification (Up/Down) and (ii) magnitude regression of standardized forward returns. Models are trained and tested in temporal order using a rolling-origin setup with non-overlapping test blocks; all results are out-of-sample. Unless noted, the classifier uses a 0.5 decision threshold, and regression is summarized by median/mean errors.

Table 1 summarizes performance metrics for the studied RF classifier.

Table 1
Performance metrics for RF classifier

	Precision	Recall	F1-score
Up	0.5842	0.5857	0.5850
Down	0.5836	0.5822	0.5829
Accuracy	0.5839	0.5839	0.5839
Macro average	0.5839	0.5839	0.5839
Weighted average	0.5839	0.5839	0.5839

The classifier shows moderate ranking ability: AUC = 0.62, i.e., a random Up case outranks a Down case ~62% of the time. Class-wise scores are tightly aligned – Up: Precision = 0.5842, Recall = 0.5857, F1-score = 0.5850; Down: Precision = 0.5836, Recall = 0.5822, F1-score = 0.5829 – indicating no material class bias at the default threshold and roughly symmetric type-I/II errors. Because the test set is essentially balanced, Accuracy = 0.5839, Macro = 0.5839, and Weighted = 0.5839 coincide. Interpreted probabilistically, the model is correct ~58.4% of the time – better than chance (50%) but economically modest without further tuning.

Performance could likely improve with (i) threshold optimization, (ii) probability calibration, and (iii) cost-sensitive training when false-positive/false-negative costs differ. Overall, the results reflect a balanced, threshold-dependent signal: the model captures useful structure, but converting ranking skill into higher decision accuracy requires careful operating-point selection and/or feature/model refinements.

Figure 1 demonstrates top-20 impurity-based feature importances for RF classifier.

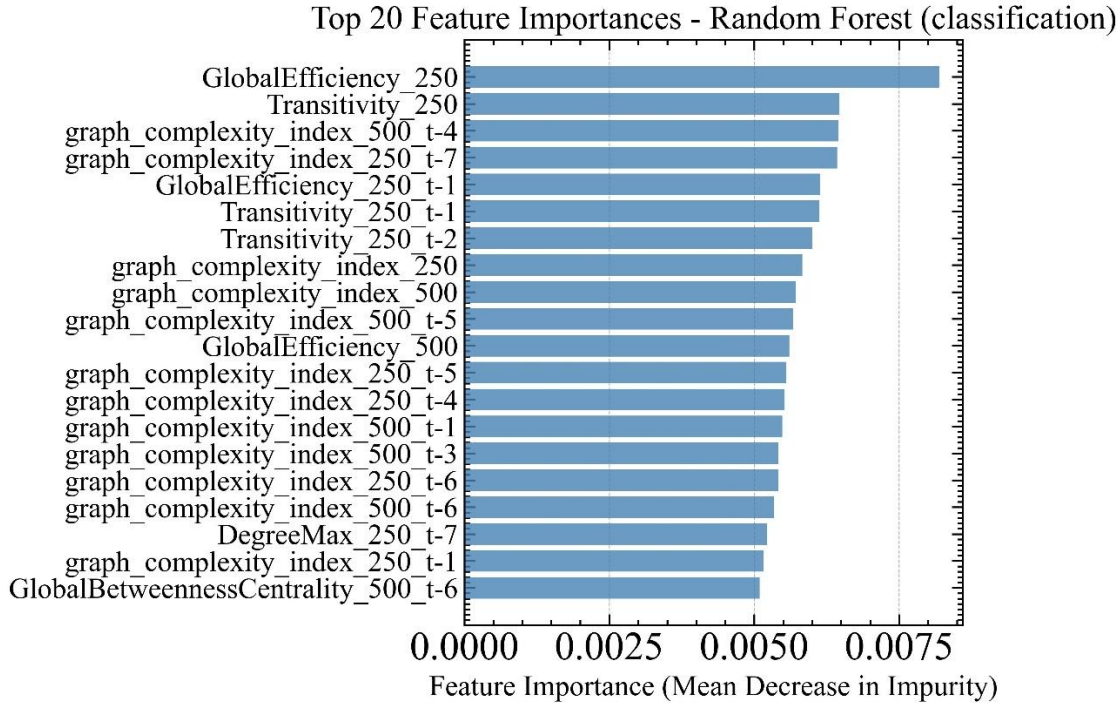

Figure 1: Top-20 impurity-based feature importances for RF classifier (7-day direction). Bars show MDI contributed by each VG-derived predictor; scores are averaged over trees and normalized. Subscripts denote the construction window (e.g., "_250", "_500").

Predictive weight is concentrated in meso-scale VG measures. Features from the 250-day window lead, 500-day contribute secondarily, and 100-day add little for a 1-week horizon. Node-centric extremes (DegreeMax) and path brokerage (GlobalBetweennessCentrality) rank near the tail, indicating reliance on global/topological organization rather than local hubs.

High efficiency (short paths) and transitivity (triadic closure) signal globally navigable yet locally cohesive structures that precede directional moves. Repeated GIC entries suggest that distance from path-like or clique-like extremes is systematically informative (or that correlated proxies capture the same regime).

MDI is relative and sensitive to feature correlation; importance can disperse across similar features and provides no direction of effect.

Features summarizing global navigability, local cohesion, and intermediate connectivity are most promising for week-ahead direction.

Table 2 demonstrates summary performance metrics obtained from the regression approach.

Table 2
Performance metrics for regression models

	R^2	MAE	MSE	RMSE
Random Forest	-0.0592	0.8758	1.2540	1.1198

Table 2 reports out-of-sample regression results: $R^2 = -0.0592$, MAE $= 0.8758$, MSE $= 1.2540$, RMSE $= 1.1198$. The negative R^2 indicates the RandomForestRegressor underperforms a mean-only baseline. Because targets are standardized, an RMSE clearly above 1 means the model does worse than a naive zero forecast. Overall, the forest captures broad, low-frequency patterns while shrinking predictions toward zero and underestimating extremes – consistent with VG features being more useful for directional ranking than for accurate magnitude prediction at a 1-week horizon.

Figure 2 represents top-20 impurity-based feature importances for the RF regressor.

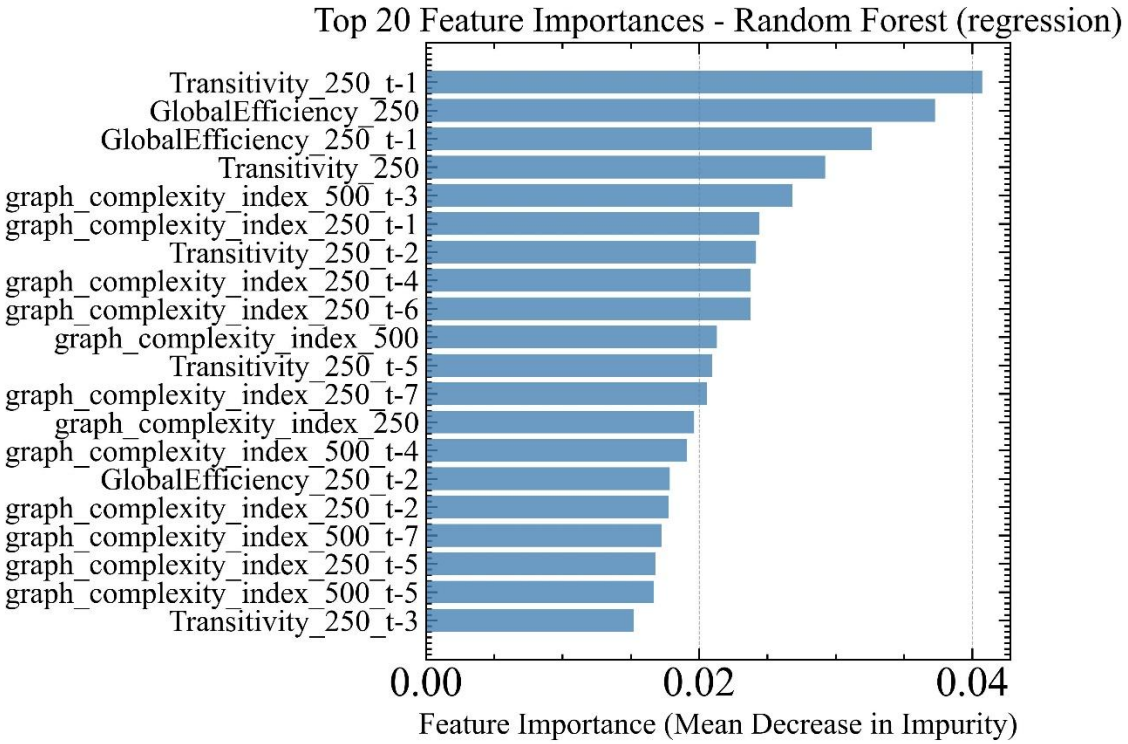

Figure 2: Top-20 impurity-based feature importances for the RF regressor (7-day ahead standardized returns). Bars report MDI for VG descriptors. Subscripts indicate the VG construction window; "$t - l$" denotes the l-day lag.

Figure 2 shows that importance is heavily concentrated in meso-scale descriptors. The top contributors are Transitivity_250 and GlobalEfficiency_250 (both current and $t - 1$), followed by a tight cluster of GIC terms at short lags. Classic node-centric extremes (maximum degree, betweenness) never enter the top-20, implying the regressor relies on distributed structure, not isolated hubs or single bottlenecks, to forecast week-ahead magnitudes.

Features built on the 250-day window dominate; 500-day GICs add secondary signal, while 100-day metrics are largely absent. This aligns with a one-year context offering the best bias–variance trade-off for a 7-day horizon – short windows are noisy, very long ones dilute regime information. The few 500-day entries suggest a slow background state still matters after medium-term structure is captured.

The mix of contemporaneous (t) and short-lag ($t-1 \ldots t-7$) versions of the same measures indicates persistence over roughly a trading week and gives the forest multiple, near-collinear split options around the forecast origin – helpful against small timing jitters.

Together, transitivity (clustering) and global efficiency (short average paths) emphasize regimes that are locally cohesive yet globally navigable – VG configurations that empirically co-move with next-week return magnitude more than node-local centralities do. Repeated GIC appearances (maximal at intermediate connectivity) further suggest that distance from trivial structures (path-like vs. clique-like) is systematically tied to return scale; the breadth of GIC lags signals either a robust link or correlated proxies of the same latent regime.

Overall, the figure indicates that medium-horizon, meso-scale organization – captured by clustering, efficiency, and spectral complexity – carries the main explanatory power for week-ahead return magnitudes, while purely local centralities are secondary. This is structurally consistent with the VG framework and guides feature engineering and robustness checks.

5. Limitations and Future Directions

This study offers useful insights but has notable constraints. First, the mutual-information prefilter may miss higher-order feature interactions, biasing selection. Second, we evaluated only one learner (RF), so we cannot judge how alternative models would capture patterns or reorder feature importance. Third, results may depend on the VG window length: we tested a few fixed sizes (100/250/500 days) and found the strongest signal near ~ 250 days, while shorter windows were noisier. Likewise, we fixed the forecast horizon at 7 trading days and targets to either a binary direction or a single 7-day return, leaving other horizons and targets unexplored.

Our evidence is also dataset-specific: we used daily S&P 500 closes only. Other indices, single stocks, and non-equity assets (bonds, commodities, FX, crypto), as well as different sampling frequencies (intraday or lower-frequency macro series), may exhibit different VG behavior; thus generalization remains an open question.

Future work should enrich the feature set with additional complexity and nonlinear descriptors – e.g., non-extensive statistics [47], entropy measures (VG or permutation entropy) [48–50], recurrence metrics [51], and fractal/scaling indicators (Hurst, fractal dimensions) [52]. Beyond RF, comparing RNN/LSTM models [53–55], GNNs that operate directly on VGs, Transformers for long-range dependence, and fuzzy or neuro-fuzzy hybrid systems [50, 56–58] would clarify accuracy–interpretability trade-offs and whether other learners surface new signal.

MDI is only one importance metric and is known to be biased toward high-cardinality features. Follow-up studies should pair it with permutation importance or SHAP, and consider wrapper or regularization-based selection (e.g., LASSO, embedded tree-ensemble methods) to validate which features are truly predictive.

Finally, broaden the task design: test multiple horizons (1-day to monthly, multi-horizon setups) and alternative targets (magnitude buckets, volatility/drawdowns, multi-output objectives). Extending across assets, markets, and frequencies – including intraday data – will determine how robust VG-based features are and where domain-specific adaptations are needed.

6. Conclusions

We pair complex network analysis with an interpretable ML approach to enhance financial time-series forecasting. Converting the S&P 500 into VG and extracting diverse topological and spectral descriptors lets us capture the market's recent structural "signature." Using RF both as predictor and as an interpretability tool, we identify which network features most influence short-horizon outcomes. Our goal is insight rather than state-of-the-art accuracy; RF's built-in importance measures make the drivers transparent. The evidence shows that combining network-based features with an interpretable ensemble offers fresh perspective on market dynamics.

The most informative predictors are meso-scale connectivity and clustering. In both classification and regression, global efficiency and transitivity computed over roughly one year consistently dominate, indicating that highly navigable, locally clustered VGs tend to precede sizable index moves. A spectral complexity score (GIC) also ranks highly, suggesting that distance from trivial structures (chains or cliques) carries signal. By contrast, node-local metrics (e.g., maximum degree, betweenness) contribute little, implying that distributed structural patterns –not isolated extremes – primarily guide forecasts.

Empirically, the RF achieves about 58% weekly direction accuracy (vs. a 50% baseline) and the regression model tracks low-frequency drift while underestimating extremes – consistent with our emphasis on understanding rather than optimizing raw performance. Crucially, the importance profiles clarify why the model forecasts as it does, a key requirement in finance.

Overall, we show that graph-derived measures of connectivity, clustering, and complexity can act as leading indicators of short-term behavior and complement traditional signals. The framework is general: it can be applied to other indices or non-financial series, or combined with more sophisticated learners while retaining interpretability. Uniting complex network science with ML thus offers a path to models that are both data-driven and structurally explainable.

Acknowledgements

The article was prepared as part of the state-funded research project 'Financial ecosystem transformation in the post-war recovery of Ukraine on the basis of resilience and sustainable development' (State Registration No. 0125U000541), conducted at Kyiv National Economic University named after Vadym Hetman.

Moreover, the authors express their profound gratitude to the Armed Forces of Ukraine, whose service and sacrifice made it possible to carry out this research during wartime.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-5 for grammar and spelling assistance. After using this tool, they reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Y. Zou, R.V. Donner, N. Marwan, J.F. Donges, J. Kurths, Complex network approaches to nonlinear time series analysis, *Phys. Rep.* 787 (2019) 1–97. doi:10.1016/j.physrep.2018.10.005.

[2] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J.C. Nuño, From time series to complex networks: The visibility graph, *Proc. Natl. Acad. Sci. USA* 105 (2008) 4972–4975. doi:10.1073/pnas.0709247105.

[3] B. Luque, L. Lacasa, F.J. Ballesteros, J. Luque, Horizontal visibility graphs: Exact results for random time series, *Phys. Rev. E* 80 (2009) 046103. doi:10.1103/PhysRevE.80.046103.

[4] X.-H. Ni, Z.-Q. Jiang, W.-X. Zhou, Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks, *Phys. Lett. A* 373 (2009) 3822–3826. doi:10.1016/j.physleta.2009.08.041.

[5] M. Stephen, C. Gu, H. Yang, Visibility graph based time series analysis, *PLOS ONE* 10 (11) (2015) e0143015. doi:10.1371/journal.pone.0143015.

[6] M.D. Vamvakaris, A.A. Pantelous, K.M. Zuev, Time series analysis of S&P 500 index: A horizontal visibility graph approach, *Physica A* 497 (2018) 41–51. doi:10.1016/j.physa.2018.01.010.

[7] C. Rudin, Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead, *Nat. Mach. Intell.* 1 (5) (2019) 206–215. doi:10.1038/s42256-019-0048-x.

[8] L. Breiman, Random Forests, *Machine Learning* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
[9] M. Gidea, Y. Katz, Topological data analysis of financial time series: Landscapes of crashes, *Physica A* 491 (2018) 820–834. doi:10.1016/j.physa.2017.09.028.
[10] V.F. Silva, R. Zhao, Time Series Analysis via Network Science, arXiv:2110.09887 (2021). URL: https://arxiv.org/abs/2110.09887.
[11] Y. Zhu, J. Wei, An entropy-based approach to the complexity of stock markets using visibility graphs, Physica A 565 (2021) 125556. doi:10.1016/j.physa.2020.125556.
[12] R. Liu, Z. Chen, Analysis of Stock Price Motion Asymmetry via Visibility Graphs, Front. Phys. 8 (2020) 539521. doi:10.3389/fphy.2020.539521.
[13] J. Iacovacci, L. Lacasa, Sequential motif profile of natural visibility graphs, Phys. Rev. E 94 (2016) 052309. doi:10.1103/PhysRevE.94.052309.
[14] T. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, Dynamics of market correlations: Taxonomy of stock market fluctuations, Phys. Rev. E 68 (2003) 056110. doi:10.1103/PhysRevE.68.056110.
[15] R.N. Mantegna, Hierarchical structure in financial markets, Eur. Phys. J. B 11 (1999) 193–197. doi:10.1007/s100510050929.
[16] D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R.N. Mantegna, E. Ben-Jacob, Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market, PLoS ONE 5 (2010) e15032. doi:10.1371/journal.pone.0015032.
[17] L. Zhao, F. Li, X. Na-na, Y. Dong, Speculation Bubbles and Network Topology in the Chinese Stock Market, Int. J. Inf. Tech. Dec. 15 (2016) 1407–1430. doi:10.1142/S0219622016500413.
[18] D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R.N. Mantegna, E. Ben-Jacob, Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market, PLoS ONE 5 (2010) e15032. doi:10.1371/journal.pone.0015032.
[19] F.X. Diebold, K. Yilmaz, On the network topology of variance decompositions: Measuring connectedness of financial firms, J. Econom. 182 (2014) 119–134. doi:10.1016/j.jeconom.2014.04.012.
[20] S.H. Kang, S.M. Nishimura, S.-M. Yoon, Network connectedness and volatility spillovers across commodity and stock markets, Energy Econ. 81 (2019) 892–907. doi:10.1016/j.eneco.2019.05.025.
[21] N. Castilho, D.Y. Kenett, E. Pinho, M. Zerenner, Forecasting Market Structure: An Edge-Prediction Approach to Correlation-Based Stock Networks, J. Netw. Theory Finance 7 (2021) 1–33. doi:10.21314/JNTF.2021.113.
[22] N. Castilho, E. Pinho, D.Y. Kenett, Link prediction in financial correlation networks for market structure forecasting, Netw. Sci. 9 (2021) 203–229. doi:10.1017/nws.2021.7.
[23] J. Wang, A. Tola, P. Luo, Portfolio Optimization Based on Network Centralities, Quant. Finance 24 (2024) 1499–1519. doi:10.1080/14697688.2023.2274416.
[24] R. Feng, J. Wang, L. Guo, STGAT: Spatial-Temporal Graph Attention Neural Network for Stock Prediction, Appl. Sci. 15 (2025) 4315. doi:10.3390/app15084315.
[25] Z.L. Wei, H.Y. An, Y. Yao, W.C. Su, G. Li, Saifullah, B.F. Sun, M.J.S. Wang, FSTGAT: Financial Spatio-Temporal Graph Attention Network for Non-Stationary Financial Systems and Its Application in Stock Price Prediction, Symmetry 17 (2025) 1344. doi:0.3390/sym17081344
[26] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, M. Sun, Graph neural networks: A review of methods and applications, *AI Open* 1 (2020) 57–81. doi:10.1016/j.aiopen.2021.01.001.
[27] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, T.-S. Chua, Temporal Relational Ranking for Stock Prediction, *ACM Trans. Inf. Syst.* 37(2) (2019) Article 16. doi:10.1145/3309547.
[28] K. Jariwala, C. Chattopadhyay, M. Patel, A Systematic Review on Graph Neural Network-based Methods for Stock Market Forecasting, *ACM Comput. Surv.* (2024). doi:10.1145/3696411.
[29] C. Agarwal, O. Queen, H. Lakkaraju, M. Zitnik, Evaluating explainability for graph neural networks, *Sci. Data* 10 (2023) 620. doi:10.1038/s41597-023-01974-x.
[30] M. Stephen, C. Gu, H. Yang, Visibility Graph Based Time Series Analysis, *PLOS ONE* 10(11) (2015) e0143015. doi:10.1371/journal.pone.0143015.

[31] V.F. Silva, M.E. Silva, P. Ribeiro, F. Silva, Time series analysis via network science: concepts and algorithms, *WIREs Data Mining and Knowledge Discovery* 11(3) (2021) e1404. doi:10.1002/widm.1404.
[32] H. Cao, T. Lin, Y. Li, H. Zhang, Stock Price Pattern Prediction Based on Complex Network and Machine Learning, *Complexity* (2019) 4132485. doi:10.1155/2019/4132485.
[33] Z. Qi, Z. Bu, X. Xiong, H. Sun, J. Cao, C. Zhang, A Stock Index Prediction Framework: Integrating Technical and Topological Mesoscale Indicators, in: *Proc. 2019 IEEE 20th Int. Conf. on Information Reuse and Integration for Data Science (IRI)*, 2019, pp. 1–8. doi:10.1109/IRI.2019.00018.
[34] J. Yao, *Classification of Stock Value Using Visibility Graph Method*, Master's Thesis, Northeastern University, Boston, MA, 2022.
[35] U. Kutluana, M. Türkmen, Classification of cardiac disorders using weighted visibility graph features from ECG signals, *Biomedical Signal Processing and Control* 87 (2024) 105420. doi:10.1016/j.bspc.2023.105420.
[36] J. Yao, *Classification of Stock Value Using Visibility Graph Method*, Master's Thesis, Northeastern University, Boston, MA, 2022.
[37] T. Wen, H. Chen, K.H. Cheong, Visibility graph for time series prediction and image classification: a review, *Nonlinear Dynamics* 110 (2022) 2979–2999. doi:10.1007/s11071-022-08002-4.
[38] H. Bocaccio, S. Iglesias-Pérez, M. Romance, R. Criado, G.B. Mindlin, On the application of Visibility Graphs in the Spectral Domain for Speaker Recognition, *arXiv* (2025). doi:10.48550/arXiv.2502.14110.
[39] M.K. Singh, S. Chaube, S. Pant, An integrated image visibility graph and topological data analysis for extracting time series features, *Decision Analytics Journal* 8 (2023) 100253. doi:10.1016/j.dajour.2023.100253.
[40] G. Campisi, S. Muzzioli, B. De Baets, A comparison of machine learning methods for predicting the direction of the US stock market on the basis of volatility indices, *International Journal of Forecasting* 40(3) (2024) 869–880. doi:10.1016/j.ijforecast.2023.07.002.
[41] A. Bielinskyi, V. Soloviev, A. Matviychuk, H. Velykoivanenko, Evaluating Feature Importance of Complex Network Measures in Random Forests, *Neuro-Fuzzy Modeling Techniques in Economics* 14 (2025) 32-80. doi:10/33111/nfmte/2025.032.
[42] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Phys. Rev. E 68 (2003) 056110. doi:10.1103/PhysRevE.68.056110.
[43] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305.
[44] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, Variable selection using random forests, *Pattern Recognit. Lett.* 31 (2010) 2225–2236. doi:10.1016/j.patrec.2010.03.014.
[45] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, A. Zeileis, Conditional variable importance for random forests, *BMC Bioinformatics* 9 (2008) 307. doi:10.1186/1471-2105-9-307.
[46] T. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: A corrected feature importance measure, *Bioinformatics* 26 (2010) 1340–1347. doi:10.1093/bioinformatics/btq134.
[47] A. O. Bielinskyi, A. V. Matviychuk, O. A. Serdyuk, S. O. Semerikov, V. V. Solovieva, V. N. Soloviev, Correlational and non-extensive nature of carbon dioxide pricing market, In: *Communications in Computer and Information Science*, Springer, 2022c, pp. 183–199. doi:10.1007/978-3-031-14841-5_12.
[48] A. Bielinskyi, O. Serdyuk, S. Semerikov, V. Soloviev, Econophysics of cryptocurrency crashes: A systematic review, *CEUR Workshop Proceedings* 3048 (2021) 31–133. URL: https://ceur-ws.org/Vol-3048/paper03.pdf.
[49] A. Bielinskyi, S. Hushko, A. Matviychuk, O. Serdyuk, S. Semerikov, V. Soloviev, Irreversibility of financial time series: A case of crisis, *CEUR Workshop Proceedings* 3048 (2021) 134–150. URL: https://ceur-ws.org/Vol-3048/paper04.pdf.

[50] A. Bielinskyi, V. Soloviev, S. Semerikov, V. Solovieva, Identifying stock market crashes by fuzzy measures of complexity, *Neuro-Fuzzy Modeling Techniques in Economics* 10 (2021) 3–45. doi:10.33111/nfmte.2021.003.

[51] A. Bielinskyi, V. Soloviev, V. Solovieva, S. Semerikov, M. Radin, Recurrence quantification analysis of energy market crises: A nonlinear approach to risk management, *CEUR Workshop Proceedings* 3465 (2023) 110–131. URL: https://ceur-ws.org/Vol-3465/paper14.pdf.

[52] A. Bielinskyi, V. Soloviev, V. Solovieva, A. Matviychuk, S. Semerikov, The analysis of multifractal cross-correlation connectedness between Bitcoin and the stock market, In: *Lecture Notes on Data Engineering and Communications Technologies*, Springer, 2023c, pp. 323–345. doi:10.1007/978-3-031-35467-0_21.

[53] T. Kmytiuk, G. Majore, Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks, *Neuro-Fuzzy Modeling Techniques in Economics* 10 (2021) 67–85. doi:10.33111/nfmte.2021.067.

[54] V. Derbentsev, A. Matviychuk, V. N. Soloviev, Forecasting of cryptocurrency prices using machine learning, In: L. Pichl, C. Eom, E. Scalas, T. Kaizoji (Eds.), *Advanced Studies of Financial Technologies and Cryptocurrency Markets*, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.

[55] V. Derbentsev, V. Bezkorovainyi, M. Silchenko, A. Hrabariev, O. Pomazun, Deep learning approach for short-term forecasting trend movement of stock indices, In: *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, IEEE, 2021, pp. 607–612. doi:10.1109/PICST54195.2021.9772235.

[56] A. Bielinskyi, V. Soloviev, V. Solovieva, H. Velykoivanenko, Fuzzy time series forecasting using semantic artificial intelligence tools, *Neuro-Fuzzy Modeling Techniques in Economics* 11 (2022) 157–198. doi:10.33111/nfmte.2022.157.

[57] A. Matviychuk, Fuzzy logic approach to identification and forecasting of financial time series using Elliott wave theory, *Fuzzy Economic Review* 11 (2) (2006) 51–68. doi:10.25102/fer.2006.02.04.

[58] V. Kabachii, R Maslii, S. Kozlovskyi, O. Dronchack, Identifying moments of deci-sion making on trade in financial time series using fuzzy cluster analysis, *Neuro-Fuzzy Modeling Techniques in Economics* 12 (2023) 175-205. doi:10.33111/nfmte.2023.175.