Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint Directional K-Shell and Feature-Guided Ranking in Complex Networks

Shima Esfandiari

sh.esfandiari.ce@gmail.com

Shiraz University

Research Article

Keywords: Directed complex networks, influential nodes, directed K-Shell decomposition, featureguided ranking, SIR model

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11703612/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Joint Directional K-Shell and Feature-Guided Ranking in Complex Networks

SHIMA ESFANDIARI[1]
[1]School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

ABSTRACT Identifying influential nodes in directed complex networks is an important problem in network science. Although K-shell decomposition can efficiently identify network cores containing highly influential nodes, its low resolution and core-like group problems limit its effectiveness for key-node identification in directed networks. Many methods address these issues by integrating K-shell with additional features for node ranking, but they do not focus on improving shell ranking. In this paper, we propose a ranking method named *Shell-based Structural-Propagation Influence* (SSPI) from a different perspective. First, a new directed K-shell representation, termed *KSOutIn*, is introduced to consider both incoming and outgoing edges. The two components of KSOutIn are then analyzed to determine their ability to provide a reliable shell ranking. Since they are insufficient when used alone or through a weighted linear combination, structural and topological features are extracted and evaluated using Linear Regression, Random Forest, and Gradient Boosting Machine models across 16 real-world and synthetic directed complex networks. PFI and SHAP are then used to identify the most important features. Finally, one of the most important features is combined with the K-shell information to construct SSPI for node ranking. Evaluation on various real-world directed networks shows that SSPI achieves the highest overall accuracy with low imprecision and time complexity.

INDEX TERMS Directed complex networks, influential nodes, directed K-Shell decomposition, feature-guided ranking, SIR model.

I. INTRODUCTION

Complex networks provide a framework for modeling real-world systems such as social, biological, financial, and transportation networks [1]–[5]. In many such systems, a small number of nodes can play a disproportionately important role in information diffusion, epidemic spreading, cascading processes, and network resilience [4], [5]. Identifying influential nodes is therefore a fundamental problem in network science [1], [6]–[8].

Node influence cannot be determined solely by the number of connections, since nodes with similar degrees may occupy very different structural positions. This has motivated a wide range of centrality and influence measures that exploit connectivity, neighborhood structure, path-based information, and core position. A central challenge is thus to determine which structural characteristics provide the most informative and discriminative description of node influence [9]–[11].

This issue is more challenging in directed networks, where incoming and outgoing edges have distinct structural meanings. Incoming links may reflect the connectivity received by a node, whereas outgoing links are directly related to its ability to reach other nodes. Ignoring edge direction or considering only one direction can therefore discard information relevant to influence assessment [12]–[14].

Among structural measures to influential-node identification, K-shell decomposition has attracted considerable attention because it characterizes the hierarchical position of nodes within the core–periphery structure of a network [15]. K-shell repeatedly removes nodes whose degree falls below a specified threshold and assigns nodes to successive shells according to the stage at which they are removed. The resulting shell index provides a compact representation of a node's coreness and its position within the network hierarchy. Its relatively low computational cost and ability to identify densely connected core regions have made K-shell an important component of many influential-node identification methods [16], [17].

Despite these advantages, the discrete nature of K-shell introduces an important resolution limitation. Multiple nodes can receive exactly the same shell index even though they may differ substantially in degree, neighborhood composition, connectivity to other shells, or actual spreading capability. Consequently, a high shell index does not necessarily imply that all nodes within that shell have comparable influence,

nor does it guarantee a complete ordering of nodes across shells. This limitation is closely related to the formation of groups of structurally similar or core-like nodes whose K-shell values are identical or insufficiently discriminative. As a result, a substantial body of research has attempted to augment K-shell with complementary structural information. For example, degree, neighborhood diversity, neighborhood distance, PageRank, and other topological characteristics have been incorporated into K-shell-based models to improve node discrimination [2], [9], [17].

The resolution and core-like-group limitations are particularly evident in directed networks. To overcome these limitations, the Multiple-Attribute Model (MAM) integrates K-shell with Degree Centrality and constructs a K-shell-based ranking formulation for directed and undirected networks [13], [14]. It also introduces three versions of directed K-shell: *KSIn*, *KSOut*, and *KSInOut* [14]. The *KSIn* and *KSOut* account for incoming and outgoing relationships, respectively. Their joint representation, denoted as *KSInOut*, preserves both directional components but also makes shell ordering a multidimensional problem. Since *KSInOut* has two components, it needs to be represented by a single value for a complete shell ordering. Moreover, the influential-node identification field requires more accurate structural features and ranking methods.

Motivated by this gap, we propose a feature-guided framework for influential-node identification in directed networks. First, we design a new version of directed K-shell, named $KSOutIn : (KSOutIn_{Out}, KSOutIn_{In})$, to preserve incoming and outgoing K-shell information simultaneously. To make it suitable for shell ranking, the relationship between its components, $KSOutIn_{Out}$ and $KSOutIn_{In}$, is analyzed. We then examine whether a linear combination of these components can provide a reliable shell ordering. Since these two components alone may be insufficient to efficiently rank directed shells, additional K-shell-based structural features are extracted.

The extracted features are systematically evaluated using three complementary learning models: Linear Regression (LR) [18], Random Forest (RF) [19], and Gradient Boosting Machine (GBM) [20]. LR provides an interpretable linear baseline, while RF and GBM capture nonlinear relationships and feature interactions. To identify informative features and reduce redundancy, model-agnostic interpretation methods, namely Permutation Feature Importance (PFI) [21] and SHAP [22], are employed. The selected features are subsequently integrated with $KSOutIn_{Out}$ to construct the final shell-ranking mechanism.

Finally, we propose SSPI and its MAM-based variants as enhanced directed K-shell ranking methods, demonstrating competitive or superior identification of influential nodes across diverse network structures and spreading regimes.

The main contributions of this study are:

- A new joint directional K-shell representation, termed *KSOutIn*, is introduced to simultaneously preserve incoming ($KSOutIn_{In}$) and outgoing ($KSOutIn_{Out}$) coreness information.
- A systematic analysis of the two directional components of *KSOutIn* is conducted to investigate their suitability for scalar shell ordering and to examine the role of their weighted linear combination.
- A feature-guided framework is developed for extracting and evaluating K-shell-based structural and topological features using Linear Regression (LR), Random Forest (RF), and Gradient Boosting Machine (GBM) across 16 directed real-world and synthetic networks.
- PFI and SHAP are employed to identify informative features and guide the selection of structural information for constructing a more discriminative shell-ranking mechanism based on $KSOutIn_{Out}$.
- The resulting SSPI method and its MAM-based variants are proposed as enhanced directed K-shell ranking methods for influential-node identification and are evaluated across diverse directed network structures and spreading conditions.

The remainder of this paper is organized as follows. Section II reviews related works. Section III presents the proposed methodology. Sections IV and V describe the evaluation setup and experimental results, respectively. Finally, Section VI concludes the paper and discusses future research directions.

II. RELATED WORKS

A. CENTRALITY-BASED APPROACHES FOR INFLUENTIAL-NODE IDENTIFICATION

Influential-node identification has traditionally relied on centrality measures that capture different structural aspects of a network. Degree Centrality characterizes local connectivity, whereas PageRank and Eigenvector Centrality incorporate the importance of neighboring nodes through recursive relationships [11], [23], [24]. LeaderRank extends the recursive propagation principle of PageRank by introducing a ground node and redistributing accumulated influence through the network. Although such global propagation-based methods can capture relationships that cannot be represented by simple local connectivity, they do not explicitly describe the hierarchical core position of a node. Moreover, their influence scores are generally obtained through recursive or iterative processes rather than through an explicit shell decomposition [25]. In directed networks, these measures can explicitly account for edge direction, but they generally emphasize particular aspects of connectivity rather than providing an explicit representation of a node's position within the network core–periphery structure.

More recent methods have combined multiple structural and propagation-related characteristics. Balanced Iterative Influence (BII) integrates local structural information, represented by the in-degree of a target node, with global influence obtained through an iterative propagation process in directed networks [12]. In a different direction, Return Random Walk Gravity (RRW) combines return random walks, effective distance, and a gravity model to incorporate local, global, and dynamic information into node influence assessment [26]. Multi-Factor Information Matrix (MFIM) further considers

node influence, neighbor influence, and mutual influence arising from feedback interactions, and is specifically formulated for directed weighted networks [27]. These approaches demonstrate the value of combining heterogeneous structural and propagation-related information; however, they do not explicitly address the problem of discriminating nodes that occupy similar K-shell positions.

B. K-SHELL-BASED APPROACHES

K-shell decomposition provides a hierarchical representation of the core–periphery structure of a network and has therefore been widely used in influential-node identification. However, its discrete shell indices may assign the same value to structurally different nodes, resulting in limited ranking resolution and the formation of core-like groups. Consequently, many studies have incorporated complementary information into K-shell-based methods.

Several approaches combine K-shell with degree or local neighborhood information. The DKS-based approach combines K-shell and Degree Centrality while incorporating self-influence and neighborhood-influence information [4]. Similarly, H-GSM and EHGSM integrates K-shell and Degree Centrality to jointly exploit local and global structural information [6], [28]. Other methods extend K-shell using richer neighborhood characteristics. The K-shell entropy approach incorporates K-shell diversity in the neighborhood together with Degree Centrality [16], while the Bubble Method extracts K-shell, K-shell diversity, and distance-related attributes within a local neighborhood to improve resolution [17]. NHM further adjusts the contributions of K-shell and degree-related features according to neighborhood structure and distance [29].

K-shell has also been combined with complementary global or recursive structural information. HNPR combines K-shell with PageRank [11] to exploit the complementary structural characteristics of the two measures [9]. CoCoFISo employs multiple graph structural properties, including K-shell and PageRank, within a multi-criteria decision-making framework and emphasizes the selection of complementary features for influential-node identification [2]. MAM explicitly addresses the low-resolution and core-like-group limitations of K-shell by integrating K-shell with Degree Centrality and constructing a K-shell-based ranking formulation [13]. Collectively, these studies demonstrate that K-shell provides valuable core-position information but often requires complementary structural attributes to achieve finer discrimination.

Recent approaches have extended this idea toward richer structural representations. GRAIN combines K-shell-based structural information with multi-order structural diversity and topological cohesion, using higher-order neighborhood information and resistance distance within an adaptive gravity-based framework [8]. Such approaches reflect a growing interest in integrating core position with local, higher-order, and propagation-related structural information. Nevertheless, their primary objective remains the construction of more informative node-level influence scores rather than the explicit ranking of multidimensional directed K-shell structures.

C. K-SHELL IN DIRECTED NETWORKS

The extension of K-shell to directed networks is challenging because incoming and outgoing connections may have different structural roles. Existing directed K-shell formulations therefore distinguish between incoming, outgoing, and joint directional connectivity [14], [30], [31]. A recent directed K-shell study introduced three variants based on incoming edges, outgoing edges, and the simultaneous consideration of both types of edges. It subsequently combined directed K-shell with directed Degree Centrality to address the resolution and core-like-group problems in directed networks [14]. These results demonstrate that the limitations associated with the discrete nature of K-shell are also relevant when directionality is explicitly considered.

When incoming and outgoing shell information are retained simultaneously, however, the shell representation is no longer a single scalar. Instead, it can be represented by a pair of directional shell components, creating a new problem: how to establish a meaningful ordering among shells with different incoming and outgoing structural positions. A simple fixed combination of these components may impose an ordering without adequately reflecting their relationship with spreading capability. Therefore, the problem is not merely to construct a directed version of K-shell, but also to determine how the resulting multidimensional shell information can be converted into a discriminative and propagation-relevant ranking.

This issue motivates the present study. Rather than introducing a predefined combination of one additional centrality measure, the proposed framework investigates a broader set of structural features associated with the directional shells. These features are evaluated using linear and nonlinear learning models, and PFI and SHAP are employed to identify characteristics that are consistently informative for shell-level spreading capability. The resulting structural evidence is then used to construct the proposed influential-node ranking mechanism. Thus, the main distinction of this study is not only the joint directional shell representation, but also the systematic analysis of which structural attributes are useful for discriminating and ranking directed shells.

III. PROPOSED METHOD

This section presents the proposed method for identifying and ranking influential nodes in directed networks. The main idea is to extend the K-shell concept to directed networks by simultaneously incorporating structural information from incoming and outgoing connections. Since directed networks contain two distinct types of relationships, considering only one direction may not adequately capture the structural characteristics of the network. Accordingly, the proposed method introduces a joint directional shell structure, referred to as *KSOutIn*, in which both incoming and outgoing connectivity

information are incorporated into the shell decomposition process.

Each KSOutIn shell is represented by an ordered pair consisting of $KSOutIn_{Out}$ and $KSOutIn_{In}$. Although this two-dimensional representation preserves complementary directional structural information, it does not directly provide a unique scalar value for ordering the shells. Since the ultimate objective is to obtain a meaningful shell ranking and subsequently rank the nodes belonging to these shells, the information contained in the two components must first be examined and then complemented with additional structural information when necessary.

The overall framework of the proposed method is illustrated in Fig. 1. The procedure begins with a directed network as the input and proceeds through several successive stages. First, the KSOutIn structure is constructed by considering both incoming and outgoing connectivity. Next, the relationship between $KSOutIn_{In}$ and $KSOutIn_{Out}$ is investigated to determine whether these two components can provide sufficient information for ranking the resulting shells. Heatmaps are employed to visualize their relationship with the spreading capability of the nodes within the corresponding shells. In addition, weighted linear combinations of $KSOutIn_{In}$ and $KSOutIn_{Out}$ are examined by introducing a weighting parameter λ to determine whether the two components can be transformed into a single scalar criterion suitable for shell ranking.

The analysis of the two KSOutIn components provides the basis for determining their suitability for shell ranking. Since the directional components alone or their weighted linear combination may not provide sufficient resolution for distinguishing the shells, additional KS-based structural features are extracted from the directed shell decomposition. These features are designed to capture complementary characteristics of the shells and are subsequently evaluated together with $KSOutIn_{Out}$ to determine their ability to distinguish and rank KSOutIn shells.

To examine the effectiveness of the extracted features from different modeling perspectives, three machine learning models are employed: Linear Regression (LR), Random Forest (RF), and Gradient Boosting Machine (GBM). LR provides a linear modeling framework, whereas RF and GBM allow nonlinear relationships and feature interactions to be considered. The models are trained using the extracted structural features and evaluated according to their ability to reproduce the shell ranking associated with spreading capability across the considered directed networks.

Feature-importance analysis is then performed using two model-agnostic methods, namely Permutation Feature Importance (PFI) and SHAP. PFI evaluates the change in model performance when the information provided by an individual feature is disrupted, while SHAP quantifies the contribution of each feature to the model output based on Shapley values. These complementary analyses are used to identify the most informative structural features and reduce the number of features included in the final ranking mechanism.

Based on the feature-importance analysis, the selected structural features are integrated with $KSOutIn_{Out}$ to construct the final shell-ranking criterion. The resulting criterion combines the directional shell position with complementary structural information and is subsequently used to rank the KSOutIn shells. The shell ranking is then used to determine the ranking of the nodes contained within the corresponding shells.

Finally, the resulting shell-ranking mechanism is incorporated into the proposed *Shell-based Structural-Propagation Influence* (SSPI) method for influential-node identification. In addition, MAM-based variants of the proposed framework are considered to provide corresponding enhanced directed K-shell ranking methods. The complete procedure is summarized in the form of pseudocode to provide a precise and reproducible description of the proposed approach.

A. KSOUTIN ALGORITHM

In this paper, we propose a directed K-shell decomposition method that incorporates both outgoing and incoming connections, termed *KSOutIn*. The method assigns each node v a two-dimensional shell index (k, l), where k and l denote the out-degree- and in-degree-based shell thresholds, respectively.

For a given pair (k, l), every remaining node satisfying $DCOut(v) \leq k$ and $DCIn(v) \leq l$ is iteratively removed and assigned to that shell. The threshold pairs are examined in lexicographic order: k advances from 1 to $\max_{DCOut}$, and for each k, the value l advances from 1 to $\max_{DCIn}$. Pairs for which no node satisfies the removal condition are skipped without assigning any shell index.

Algorithm 1 presents an efficient implementation of this procedure using two degree-bucket structures, B_{out} and B_{in}, indexed by the current out-degree and in-degree of unremoved nodes, respectively. Only unremoved nodes are maintained in B_{out}, whereas nodes that have become eligible with respect to the current threshold k (i.e., $DCOut(v) \leq k$) are maintained in B_{in}. When a node is removed, the degrees of its neighbors are decremented through the adjacency lists and the affected nodes are moved to their updated degree buckets. A node whose current out-degree drops to at most k is immediately activated, while an active node whose in-degree drops to at most l is immediately scheduled for removal. Empty (k, l) pairs are bypassed by advancing directly to the next nonempty bucket, so no redundant scans occur. The resulting decomposition is identical to the straightforward iterative procedure because the same removal condition and threshold order are preserved throughout.

Unlike conventional one-dimensional K-shell decomposition, each node in the proposed method is therefore characterized along two structural dimensions: its out-degree and in-degree. The resulting two-dimensional shell index provides a more detailed representation of the structural position of nodes in directed networks, and subsequently serves as the basis for the structural feature extraction and shell-ranking stages of the proposed method.

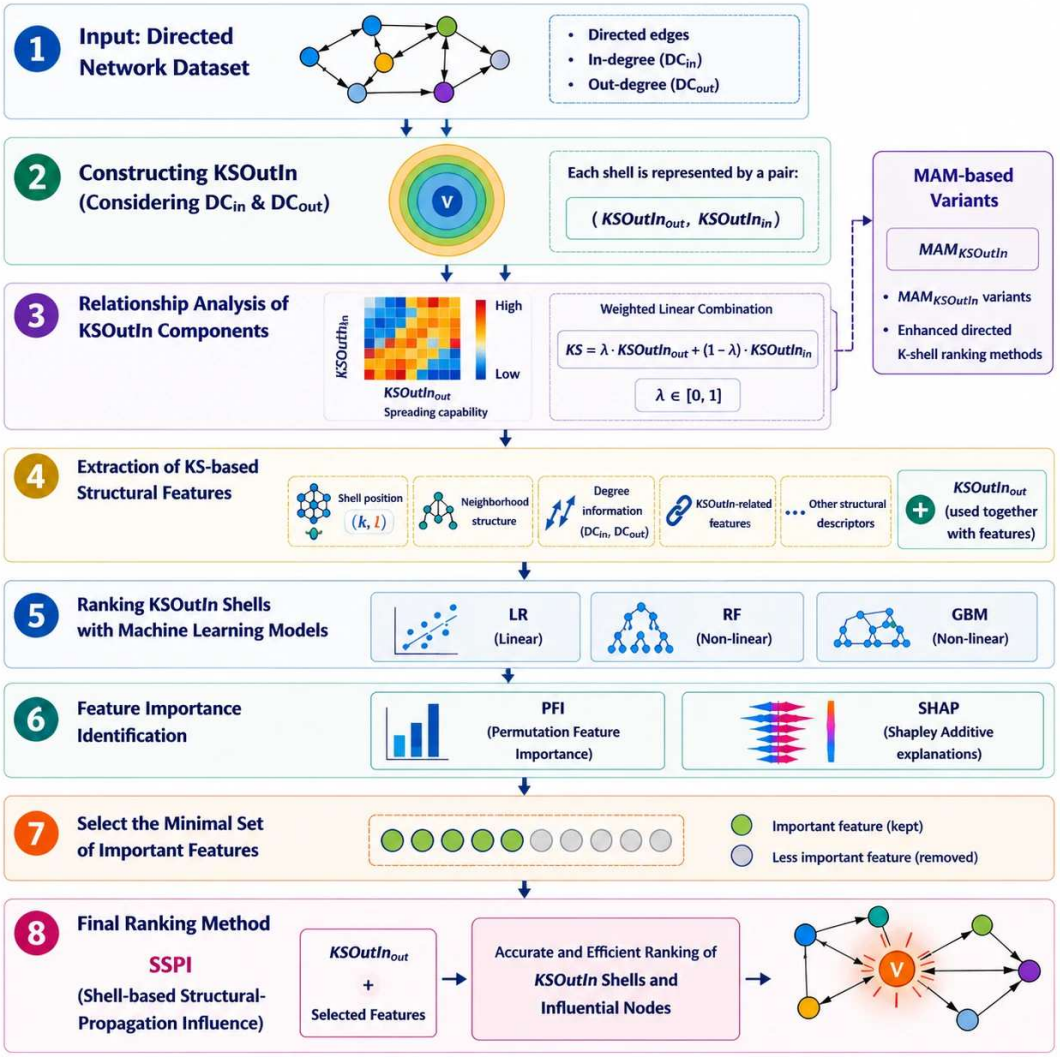


FIGURE 1. **Overall framework of the proposed method for identifying and ranking influential nodes in directed networks.**

Time Complexity

The bucket-based implementation achieves an overall time complexity of $O(n+m)$. The initial computation of $DCOut$ and $DCIn$ and the construction of the adjacency lists each require $O(n+m)$ time. During the decomposition, every node is activated and removed at most once. Every directed edge (u,w) is examined exactly once, when one of its endpoints is removed, so each degree decrement and the corresponding bucket update require $O(1)$ amortized time. Threshold advancement is driven by the bucket structure: a value of k or l for which no node satisfies the removal condition is bypassed in $O(1)$ by moving to the next nonempty bucket, so empty (k,l) pairs incur no per-node cost. Summing over all nodes, edges, and threshold steps, the total cost of node processing, edge updates, bucket operations, and threshold advancement is $O(n+m)$.

The decomposition produced by Algorithm 1 is identical to the straightforward iterative procedure, because the bucket structure preserves both the removal condition ($DCOut(v) \leq k$ and $DCIn(v) \leq l$) and the lexicographic threshold order; it only eliminates redundant scans over nodes and empty threshold pairs. For sparse directed networks in which $m = O(n)$, the complexity reduces to $O(n)$.

B. SCALARIZATION OF THE KSOUTIN DECOMPOSITION

The KSOutIn decomposition assigns each node an ordered pair consisting of $KSOutIn_{In}$ and $KSOutIn_{Out}$. Since this two-dimensional representation does not directly provide a unique scalar ordering of the shells, a weighted combination of the two components is introduced:

$$KSOutIn = \lambda * KSOutIn_{Out} + (1-\lambda) * KSOutIn_{In} \quad (1)$$

where λ determines the relative contribution of $KSOutIn_{Out}$ to the resulting scalar index. The appropriate value of λ is determined empirically by examining its relationship with the actual spreading capability of the nodes. The results of this analysis are presented in the evaluation section and are

Algorithm 1 KSOutIn: Efficient decomposition using degree buckets

Require: Directed graph $G = (V, E)$
Ensure: KSOutIn shell index (k, l) for every node $v \in V$

```
Compute DCOut[v] and DCIn[v] for all v ∈ V
Build bucket arrays B_out[0, ..., n−1] and B_in[0, ..., n−1]
Insert every node v into B_out[DCOut[v]]
k ← 1
while unremoved nodes exist do
    Activate all nodes with DCOut ≤ k; insert them into B_in
    l ← 1
    while active nodes exist do
        Advance l to the smallest nonempty bucket in B_in
        Remove all active nodes with DCIn ≤ l; assign shell index (k, l) to each
        for all incident edges (u, w) of each removed node u do
            Decrement DCOut[w] or DCIn[w] accordingly
            Move w to its updated bucket
            if DCOut[w] ≤ k and w is not yet active then
                Activate w; insert into B_in[DCIn[w]]
            else if w is active and DCIn[w] ≤ l then
                Schedule w for immediate removal
            end if
        end for
    end while
    Advance k to the next nonempty bucket in B_out
end while
```

subsequently used to motivate the extraction of additional structural features for the ranking models.

C. PROPOSED KS-BASED FEATURES FOR SHELL RANKING

Since $KSOutIn_{Out}$ alone does not fully characterize the structural complexity of a shell, we propose a set of KS-based features. These features are designed to capture the internal structure of each shell, its connectivity with the rest of the network, and its relative position within the *KSOutIn* decomposition process. The primary objective is to provide complementary information to machine learning models, thereby identifying which specific characteristics most significantly influence a shell's structural importance and spreading capability.

Let S denote a shell comprising N_S nodes. The features extracted for each shell are defined as follows:

1) Internal Connectivity Normalized by KS

This feature quantifies the internal connectivity of a shell relative to the *KS* values of its constituent nodes. A shell with high internal connectivity typically signifies a dense structural organization; thus, this feature serves as an indicator of the shell's structural significance.

$$f_1_internalConn = \frac{2E_{\text{in}}(S)}{\sum_{n\in S}\left[KSOutIn_{Out}(n)^2 + KSOutIn_{In}(n)^2\right]}, \tag{2}$$

where $E_{\text{in}}(S)$ represents the number of internal edges within shell S. A higher value of f_1*_internalConn* signifies greater internal connectivity relative to the *KS* values of the shell's nodes.

2) Boundary Out-Edges per Node

This feature measures the extent to which a shell connects to external nodes via outgoing edges. Given that the capacity to propagate influence across a directed network is largely dependent on outgoing connections, this metric is crucial for identifying shells with superior spreading capabilities.

$$f_2_boundaryOutPerNode = \frac{E_{\text{Out}}(S)}{N_S}, \tag{3}$$

where $E_{\text{Out}}(S)$ denotes the count of edges directed from nodes within the shell to nodes outside the shell. A higher value of f_2*_boundaryOutPerNode* indicates a stronger outgoing influence from the shell to the remainder of the network.

3) Mean Out-Degree

The mean out-degree represents the average number of direct outgoing connections for nodes within a shell. Since these edges provide direct pathways for information or influence propagation, this feature is intrinsically linked to the shell's potential for spreading.

$$f_3_meanDCOut = \frac{1}{N_S}\sum_{n\in S} DC_{\text{Out}}(n), \tag{4}$$

where $DC_{\text{Out}}(n)$ is the out-degree of node n. Higher values of f_3*_meanDCOut* suggest a greater outgoing capacity for the shell's nodes.

4) Mean In-Degree

This feature assesses the average incoming connectivity of nodes within a shell. While in-degree does not directly measure spreading capability, it indicates the extent to which a shell acts as a recipient of information flow, thereby reflecting its position within the broader network structure.

$$f_4_meanDCIn = \frac{1}{N_S}\sum_{n\in S} DC_{\text{in}}(n), \tag{5}$$

where $DC_{\text{in}}(n)$ denotes the in-degree of node n. A higher value of f_4*_meanDCIn* implies that the nodes in the shell receive a higher volume of incoming connections.

5) Iteration Factor

The *KSOutIn* decomposition process is sequential; the order in which nodes are pruned reveals their structural centrality. This feature captures the stage at which the nodes of a shell

are removed, thereby distinguishing shells based on their topological hierarchy.

$$f_5_iterationFactor = \max_{n \in S} Iteration(n), \tag{6}$$

where $Iteration(n)$ denotes the removal iteration of node n. This value reflects the latest removal stage among all nodes in the shell, effectively differentiating shells with similar *KS* values that emerge at distinct decomposition stages.

6) Diameter of the Largest Connected Component

This feature characterizes the spatial extent and topological organization of a shell. By treating the shell as an undirected graph, we compute the diameter of its largest connected component to capture the structural distance between nodes.

$$f_6_diameter = Diameter(C_{\max}(S)), \tag{7}$$

where $C_{\max}(S)$ represents the largest connected component of the shell after ignoring edge directions. A larger f_6*_diameter* indicates a more extended structure with greater topological distances between nodes.

7) Shell $KSOutIn_{Out}$

As demonstrated in the heatmap analysis (Section V-A), there is a strong correlation between $KSOutIn_{Out}$ and spreading capability. Consequently, this feature is included as a primary structural descriptor.

$$f_7_KSOutIn_{Out} = KSOutIn_{Out}(S), \tag{8}$$

where $KSOutIn_{Out}(S)$ is the $KSOutIn_{Out}$ value associated with shell S, representing its structural position derived from out-degree decomposition.

8) Shell $KSOutIn_{In}$

To evaluate the contribution of in-degree-based structural information, we incorporate the $KSOutIn_{In}$ value of each shell as an independent feature.

$$f_8_KSOutIn_{In} = KSOutIn_{In}(S). \tag{9}$$

This feature provides the model with the in-degree-based structural context of the shell, allowing for an examination of the relative importance of $KSOutIn_{In}$ versus $KSOutIn_{Out}$.

9) Maximum $KSOutIn_{Out}$ in the Network

The significance of a shell's $KSOutIn_{Out}$ value may be relative to the network's overall scale. By including the global maximum $KSOutIn_{Out}$ value, the model can normalize and evaluate a shell's relative importance within the network's hierarchy.

$$f_9_maxKSOutIn_{Out} = \max_{n \in G} KSOutIn_{Out}(n). \tag{10}$$

10) Maximum $KSOutIn_{In}$ in the Network

Similarly, we incorporate the maximum $KSOutIn_{In}$ value observed in the entire network to characterize the overall scale of the in-degree-based decomposition.

$$f_{10}_maxKSOutIn_{In} = \max_{n \in G} KSOutIn_{In}(n). \tag{11}$$

11) Total Number of Shells

The number of shells generated by the *KSOutIn* decomposition often reflects the structural differentiation of a network. This global feature accounts for the total number of unique shells, providing context for the model to interpret the granularity of the structural decomposition.

$$f_{11}_totalShells = |S_G|, \tag{12}$$

where S_G denotes the set of unique shells identified in network G.

In summary, these proposed features capture multifaceted aspects of shell structure. Features f_1*_internalConn* and f_6*_diameter* describe the topological and structural properties; f_2*_boundaryOutPerNode* through f_4*_meanDCIn* characterize directed connectivity; f_5*_iterationFactor* captures temporal information related to the removal process; f_7*_*$KSOutIn_{Out}$ and f_8*_*$KSOutIn_{In}$ provide direct KS-based metrics; and f_9*_*$maxKSOutIn_{Out}$ through f_{11}*_totalShells* represent global network characteristics. The integrated use of these features enables a more robust assessment of shell importance compared to relying solely on $KSOutIn_{Out}$.

D. PROPOSED METHOD: SHELL-BASED STRUCTURAL-PROPAGATION INFLUENCE (SSPI)

Based on the results obtained from the linear and nonlinear models, as well as the feature-importance analysis presented in Section V-D, the *Shell-based Structural-Propagation Influence* (SSPI) criterion is proposed for identifying and ranking influential nodes in directed networks. The design of SSPI is guided by these feature analysis and combines shell-level structural information with the propagation-related information of the nodes reachable from each source node.

SSPI consists of two main components. The first component represents the structural position of a node through the shell to which it belongs. The second component captures the outgoing connectivity of the nodes that can be reached from the source node through directed paths. This combination enables the influence of a node to be evaluated according to both its structural position within the network and the outgoing capacity of its reachable neighborhood.

1) Structural Component

The structural component is based on the proposed KSOutIn decomposition. According to the feature-importance analysis, $KSOutIn_{Out}$ and the mean out-degree of the nodes within each shell, denoted by f_3*_meanDCOut*, are among the informative features for shell ranking. Based on the results of

Scenario I, these two features are selected to construct the structural component of SSPI.

Let S_i denote the shell containing node i. The *K-Shell Extended* (KSE) value of node i is defined as

$$KSE(i) = \sqrt{KSOutIn_{Out}(S_i) \times f_3_meanDCOut(S_i)}, \quad (13)$$

where $KSOutIn_{Out}(S_i)$ denotes the common $KSOutIn_{Out}$ value of shell S_i, and $f_3_meanDCOut(S_i)$ denotes the mean out-degree of the nodes belonging to that shell. Since both quantities are defined at the shell level, all nodes within the same shell have the same KSE value. Therefore, KSE determines the structural component of the ranking, whereas the distinction between nodes belonging to the same shell is provided by the propagation-based component.

2) Propagation Component

The propagation component considers the nodes reachable from node i through directed paths. For each reachable node j, its contribution is determined by its outgoing degree and its directed distance from node i. To reduce the influence of distant nodes, an inverse-square distance weight is applied. The resulting SSPI criterion is defined as

$$SSPI(i,r) = KSE(i) \times \sum_{\substack{j \\ 1 \le d(i,j) \le r}} \frac{DC_{\mathrm{Out}}(j)}{d(i,j)^2}, \quad (14)$$

where $d(i,j)$ denotes the length of the shortest directed path from node i to node j, and r denotes the maximum distance considered in the calculation. Thus, a node obtains a higher SSPI value when it belongs to a shell with a larger structural component and can reach nodes with higher outgoing degree within the specified distance. The factor $1/d(i,j)^2$ progressively reduces the contribution of reachable nodes as their distance from the source node increases.

3) Determination of the Radius

The radius r determines the range over which the propagation component is calculated. A larger radius includes a greater number of reachable nodes but may increase the computational cost, whereas a smaller radius restricts the analysis to a more local propagation environment. Therefore, the radius is selected according to the characteristics of each network rather than being fixed to a single value for all networks. In this study, the radius is restricted to the range $1 \le r \le 3$.

The radius is determined using two network characteristics: the average degree and the threshold spreading probability. When the average degree satisfies the specified threshold condition, $r = 1$ is selected. If this condition is not satisfied but the threshold spreading probability meets its corresponding condition, $r = 3$ is selected. In all remaining cases, $r = 2$ is used as an intermediate value.

Applying these criteria to the considered real-world directed networks results in the following radius assignments. The Bison, Cintestinalis, DomStrauss, and Openflight networks are assigned $r = 1$. The Uni_email and Stelzl networks are assigned $r = 2$. Finally, the Physician and Faa_routes networks are assigned $r = 3$.

The threshold values used to determine the radius are specified together with the experimental settings in the evaluation section. This arrangement keeps the proposed SSPI formulation independent of a particular threshold setting while ensuring that the radius-selection procedure remains reproducible.

Overall, SSPI combines shell-level structural information with propagation-related information in a single ranking criterion. The structural component is determined by $KSOutIn_{Out}$ and $f_3_meanDCOut$, whereas the propagation component incorporates the DC_{Out} values of reachable nodes and their directed distances from the source node. The radius r controls the range of the propagation analysis according to the characteristics of each network. The resulting SSPI value is used to rank nodes according to their structural position and propagation-related connectivity.

E. ADDITIONAL COMPARATIVE VARIANTS

In addition to the proposed SSPI criterion, several variants are considered to provide a systematic comparison with existing approaches and to examine the contribution of the proposed KSOutIn representation. In [14], three directed variants of the k-shell method were introduced, and the corresponding MAM method for directed networks employs these variants within its formulation, as defined in Eq. 17. In the present study, the KSOutIn in Eq. 1 is replaced by the KS index given in Eqs. 15- 17. Moreover, the degree component is instantiated using either the out-degree centrality (*DCOut*) or the in-degree centrality (*DCIn*). Accordingly, two additional variants, denoted by *MAM_KSOutIn_DCOut* and *MAM_KSOutIn_DCIn*, are constructed. These variants are then evaluated together with the other degree–k-shell combinations considered within the MAM framework, allowing the effect of the proposed KSOutIn representation to be examined under different degree configurations.

$$SI(i) = \frac{DC_i}{KS_i} \quad (15)$$

$$NI(i) = \frac{\sum j \in \phi_{i,1} DC_j}{KS_i^2} \quad (16)$$

$$MAM(i) = \alpha KS_i \times \beta \sum_{j \in \phi_{i,2}} \frac{\frac{NI(j)}{max(NI)} + \frac{SI(j)}{max(SI)} + \frac{KS_j}{max(KS)}}{d_{ij}^2}, \quad \alpha, \beta = 1 \quad (17)$$

Furthermore, an additional ablation variant is introduced to investigate the contribution of the shell-level mean out-degree used in the KSE component of the proposed method. Specifically, the corresponding feature is excluded from the

KSE formulation, and the resulting criterion is referred to as *SSPID*. For a node i and a radius r, SSPID is defined as

$$SSPID = \sqrt{KSOutIn_{Out}(i)} \sum_{j \neq i \ d(i,j) \leq r} \frac{DCOut(j)}{d(i,j)^2}, \quad (18)$$

where $d(i,j)$ denotes the shortest directed-path distance from node i to node j, and r represents the maximum radius considered in the network. The formulation preserves the structural contribution of $KSOutIn_{Out}$ while quantifying the propagation potential of the nodes reached from i through their out-degree centrality. The inverse-square distance term further reduces the contribution of nodes located farther from the source node. Therefore, comparison between SSPID and the complete SSPI formulation provides a direct ablation analysis of the role of the shell-level mean out-degree in enhancing the proposed influence measure.

IV. EVALUATION SETUP

This section introduces the datasets and evaluation metrics. The datasets comprise of 16 directed networks consist of 8 real-world and 8 synthetic networks. Evaluation metrics include the accuracy measured by Kendall's Tau correlation coefficient [32], precision via the Imprecision function [15], resolution using the Monotonicity index [36] and time complexity.

A. REAL-WORLD DIRECTED NETWORKS

To evaluate the performance of the proposed method on real-world networks, a collection of directed networks with different characteristics and structural properties was selected from the Netzschleuder database.

The selected networks cover a variety of domains, including air transportation, social relationships, biological systems, protein interactions, electronic communication, and behavioral interactions. The Physician network represents trust and consultation relationships among physician in four cities in the Midwestern United States in 1966, where the direction of an edge indicates trust or a consultation request from one physician to another. The Cintestinalis network represents the complete synaptic connectivity structure of the brain of a larval *Ciona intestinalis*, where the direction of an edge is from a presynaptic cell to a postsynaptic cell. The Faa-routes network represents preferred flight routes recorded by the Federal Aviation Administration (FAA), where nodes correspond to airports or service centers and directed edges represent preferred routes between them.

The Stelzl network models protein–protein interactions among human proteins, while the Uni-email network represents email exchanges among members of the University of Rovira i Virgili. The Bison network describes dominance relationships among a group of American bison, whereas the DomStrauss network represents winner–loser relationships in aggressive interactions among animals. Finally, the Openflight network contains regular commercial flights between airports around the world, with the direction of an edge indicating the origin and destination of a flight.

All of these networks are treated as directed graphs. Therefore, unlike undirected networks, edge directions play an essential role in interpreting relationships as well as in calculating and ranking node importance. To provide an overview of the structural differences among the datasets, the topological characteristics of the networks, including the number of nodes (n), the number of directed edges (m), the average degree ($\langle k \rangle$), the clustering coefficient (C), and the degree assortativity coefficient (r), are reported in Table 1.

As shown in Table 1, the network size ranges from the small Bison network with 26 nodes to the large Openflight network with 3214 nodes. The number of edges varies from 222 to 18859. This diversity in network size, density, and structural characteristics provides an appropriate setting for evaluating effective node-ranking methods under different conditions and across real-world networks with heterogeneous structures.

TABLE 1. Statistical Properties of the 8 Directed Real-World Networks

Directed Network	n	m	$\langle k \rangle$	C	r	β_{th}
Physician	241	1098	7.7	0.310	-0.055	0.21
Cintestinalis	205	2624	25.6	0.350	0.135	0.06
Faa-routes	1226	2410	3.9	0.068	-0.013	0.26
Stelzl	1706	3191	3.7	0.006	-0.189	0.06
Uni-email	1133	5452	9.6	0.220	0.078	0.05
Bison	26	222	17.1	0.813	-0.104	0.08
DomStrauss	150	1882	25.1	0.651	-0.035	0.07
Openflight	3214	18859	11.7	0.492	-0.017	0.02

B. SYNTHETIC DIRECTED NETWORKS

1) Barabási–Albert (BA) Networks

To evaluate the performance of the proposed method on networks with scale-free structures and heterogeneous degree distributions, the Barabási–Albert (BA) model was employed [33]. This model is based on the principle of preferential attachment, whereby newly added nodes are more likely to connect to nodes that already have higher degrees. As a result, the generated network contains a limited number of high-degree nodes (hubs) and a large number of low-degree nodes.

In this study, the number of nodes in the network was set to 1000, and three different values of the parameter m, namely 2, 4, and 6, were considered. The parameter m specifies the number of edges created by each newly added node. Therefore, increasing m results in a larger number of edges and a higher average degree.

According to the generated datasets, the numbers of edges for $m = 2$, $m = 4$, and $m = 6$ were 1996, 3984, and 5964, respectively. The corresponding average degrees were 3.99, 7.96, and 11.93, respectively.

2) Lancichinetti–Fortunato–Radicchi (LFR) Networks

To investigate the performance of the ranking method on networks with community structures and internal organization, the Lancichinetti–Fortunato–Radicchi (LFR) model was employed [34]. This model enables the generation of networks

with heterogeneous degree distributions and community-based structures while controlling properties such as node-degree distributions, community sizes, and the degree of mixing between communities.

In this study, the number of nodes was set to 1000, the community-size distribution parameter was set to $\tau_2 = 1.5$, the minimum community size was set to 50, and the target average degree was set to 10. To investigate the effect of network structure, different values of τ_1, namely 2, 2.5, and 3, and different values of μ, namely 0.1, 0.2, and 0.4, were considered.

The parameter τ_1 controls the shape of the degree distribution, whereas μ represents the mixing level of nodes between different communities. Therefore, increasing μ increases the proportion of inter-community connections and reduces the structural separation between communities.

In the generated networks, the number of edges ranged from 5876 to 7198, while the average degree ranged from 11.75 to 14.40. The clustering coefficient also varied from approximately 0.108 to 0.486 depending on the network parameters, indicating substantial differences in the local structure of the generated networks.

C. GENERATION OF DIRECTED NETWORKS

Since the objective of this study is to investigate effective node-ranking methods in directed networks, the BA and LFR networks were initially generated as undirected networks, after which directions were assigned to their edges.

In the primary setting, the direction of each edge was assigned randomly in order to construct a directed network while preserving the original topology. In addition, a degree-based edge orientation scheme was considered to investigate the robustness of the methods to changes in the edge-orientation pattern.

In this setting, using a probability of 0.7, edge directions were assigned such that higher-degree nodes predominantly act as the sources of edges and consequently have a larger number of outgoing edges. In other words, the edge orientation is associated with the degree structure of the network, allowing the ranking methods to be evaluated under a non-random directional distribution.

For both models, the number of nodes was set to 1000, and the structural properties of the networks were calculated and recorded after generation. These properties included the number of nodes and edges, network density, the mean and standard deviation of the in-degree and out-degree, maximum degree, the number of weakly and strongly connected components, the clustering coefficient, and the characteristics of the largest component (Table. 2 and 3).

For the LFR networks, the number of communities and the modularity measure were also calculated based on the initial undirected topology. It should be noted that orienting the edges does not change the number of edges or the underlying undirected topology. Therefore, the reported values of the number of edges, average degree, clustering coefficient, and degree assortativity represent the properties of the original undirected topology, whereas properties such as in-degree and out-degree are affected by the edge-orientation process.

TABLE 2. Statistical Properties of the Directed BA Networks with 1000 nodes

Network	m	$\langle k \rangle$	C	r	β_{th}
BA_m2	1996	3.99	0.031	-0.094	18
BA_m4	3984	7.96	0.036	-0.064	0.12
BA_m6	5964	11.93	0.039	-0.037	0.09

TABLE 3. Statistical Properties of the Directed LFR Networks with 1000 nodes

Network	m	$\langle k \rangle$	C	r	β_{th}	τ_1	μ
LFR1	7099	14.20	0.486	-0.305	0.02	2.0	0.2
LFR2	5978	11.90	0.260	-0.165	0.10	2.5	0.1
LFR3	7198	14.40	0.228	-0.182	0.04	2.5	0.2
LFR4	6637	13.27	0.100	-0.158	0.06	2.5	0.4
LFR5	5876	11.75	0.108	-0.0974	0.11	3.0	0.2

D. EVALUATION METRICS

1) Kendall's Tau Correlation Coefficient (Accuracy)

Kendall's Tau is used to measure the agreement between the ranking produced by a method and the SIR-based ranking [35]. For each pair of nodes, their orders are classified as concordant or discordant. When tied ranks occur, Kendall's Tau-b is used:

$$\tau = \frac{C - D}{\sqrt{(C + D + T)(C + D + U)}}, \tag{19}$$

where C and D are the numbers of concordant and discordant pairs, and T and U denote tied pairs in the two rankings, respectively. $\tau \in [-1, 1]$, where values closer to 1 indicate stronger agreement with the SIR ranking. Since this metric considers all nodes, precision and the top-10 influential nodes are also evaluated.

2) Imprecision

The imprecision function [15] evaluates how effectively a method identifies the most influential fraction p of nodes:

$$I(p) = 1 - \frac{\langle M_{\text{eff}}(p) \rangle}{\langle M_{\text{SIR}}(p) \rangle}, \tag{20}$$

where pN is the number of selected top-ranked nodes, $\langle M_{\text{eff}}(p) \rangle$ is their average spreading capability, and $\langle M_{\text{SIR}}(p) \rangle$ is the corresponding value for the top pN nodes identified by SIR. Lower imprecision indicates better identification of influential nodes.

3) Monotonicity Index (Resolution)

The monotonicity index measures the resolution of a ranking by quantifying tied ranks [36]:

$$M(R) = \left[1 - \frac{\sum_{r \in R} f_r (f_r - 1)}{N(N-1)}\right]^2, \tag{21}$$

where N is the number of nodes and f_r is the number of nodes assigned to rank r. The index ranges from 0 to 1; higher values indicate fewer ties and finer ranking resolution.

4) Time Complexity

Time complexity evaluates the computational efficiency of ranking methods, particularly as network size increases. Lower computational cost is desirable when maintaining adequate ranking accuracy, especially for large-scale and dynamic networks. Therefore, time complexity is considered alongside ranking performance to assess the practical applicability of different methods.

V. EVALUATION RESULTS

This section presents a comprehensive evaluation of the proposed framework and the developed directed K-shell-based ranking methods. First, the relationship between $KSOutIn_{Out}$ and $KSOutIn_{In}$ is investigated (Section V-A), and the effect of the weighting parameter λ on shell ordering is examined (Section V-B). Next, the ability of the proposed framework to address core-like groups and resolution limitations is evaluated (Section V-C), followed by an analysis of feature importance and the performance of LR, RF, and GBM under different training and testing scenarios (Section V-D). The ranking performance of the proposed methods is then assessed against existing directed ranking methods using Kendall's Tau (V-E) and Imprecision based on SIR spreading capability across real-world directed networks (V-F). Finally, the computational complexity of the proposed methods is analyzed and compared with that of the benchmark approaches (V-G).

A. ANALYSIS OF THE RELATIONSHIP BETWEEN $KSOUTIN_{OUT}$, $KSOUTIN_{IN}$

As discussed previously, the output of the KSOutIn algorithm for each node is an ordered pair consisting of $KSOutIn_{Out}$ and $KSOutIn_{In}$. Since the ultimate objective is to rank the shells and, consequently, assign a meaningful rank to the nodes belonging to each shell, this two-dimensional representation needs to be transformed into a single scalar value. Therefore, this section investigates the relationship between the two KSOutIn components and the spreading capability of the nodes within their corresponding shells, with the aim of determining whether an appropriate combination of these components can be used as a basis for shell ranking.

Regarding the relationship between the in- and out-based KS components, the study by Luo et al. [31] provides a relevant perspective. In that study, an algorithm for identifying (k)-cores in directed networks is proposed in which the in-degree and out-degree of nodes are considered simultaneously to identify densely connected subgraphs. Within the (l, m)-core framework, the authors show that the point corresponding to the densest subgraph can be identified by maximizing the product of l and m. In other words, the quantity $l \times m$ is employed as a criterion for determining the optimal subgraph.

Motivated by this idea, and considering that k-shell-based approaches are also intended to identify structurally important nodes located in central or densely connected regions of a network, this study investigates whether the combination of $KSOutIn_{Out}$ and $KSOutIn_{In}$ can similarly be used to identify shells and nodes with greater spreading capability. More specifically, the objective is to determine whether there exists an appropriate combination of these two components that can transform the two-dimensional KSOutIn representation into a meaningful scalar measure for ranking the shells.

To investigate this issue, the KSOutIn algorithm was first applied to all considered directed networks. Subsequently, for each shell, the average spreading capability of the nodes belonging to that shell was calculated. The resulting distributions are visualized as heatmaps in Fig. 2. In these heatmaps, the horizontal axis represents $KSOutIn_{Out}$, the vertical axis represents $KSOutIn_{In}$, and the color intensity of each cell indicates the average spreading capability of the nodes belonging to the corresponding shell. As illustrated by the heatmaps, relatively strong relationships between $KSOutIn_{In}$ and $KSOutIn_{Out}$ can be observed in the Uni_email, Openflight, and Stelzl networks. Accordingly, the corresponding heatmaps exhibit a predominantly diagonal pattern, indicating that larger values of one KS component are generally accompanied by larger values of the other. As a result, distinguishing the independent contribution of the two components is more difficult in these networks.

Nevertheless, across the networks considered in this study, the highest spreading capabilities are generally observed in regions associated with larger values of $KSOutIn_{Out}$. In contrast, an increase in $KSOutIn_{In}$ does not consistently correspond to an increase in spreading capability. For example, in the Physician network, the shells $(2, 6)$ and $(2, 7)$ differ in their KSOutIn components, while the average spreading capability of the nodes in shell $(2, 6)$ is higher than that of the nodes in shell $(2, 7)$. This observation indicates that simultaneously increasing the in- and out-based KS components does not necessarily lead to higher spreading capability and that the two components do not contribute equally to the spreading process. In particular, the heatmap analysis suggests that, for the networks considered here, $KSOutIn_{Out}$ has a stronger association with spreading capability and may therefore provide more discriminative information for shell ranking than $KSOutIn_{In}$.

B. LINEAR COMBINATION OF KSOUTIN COMPONENTS USING THE WEIGHTING PARAMETER λ

The heatmap analysis indicates that $KSOutIn_{Out}$ exhibits a stronger association with the spreading capability of nodes than $KSOutIn_{In}$. To further quantify the relative contribution of these two components to shell ranking, a linear combination of $KSOutIn_{In}$ and $KSOutIn_{Out}$ is considered, as defined in Eq. 1 in the proposed method.

For different values of λ in the range $[-1, 1]$, Kendall's Tau (τ) is calculated between the ranking obtained from the combined criterion, KS_λ, and the actual spreading capability of the nodes. The objective of this analysis is to identify the value of λ that yields the strongest correspondence between the ranking induced by KS_λ and the observed spreading capability. As shown in Fig. 3, in the Physician, Bison, Faa_routes, Stelzl, and Openflight networks, Kendall's Tau increases as

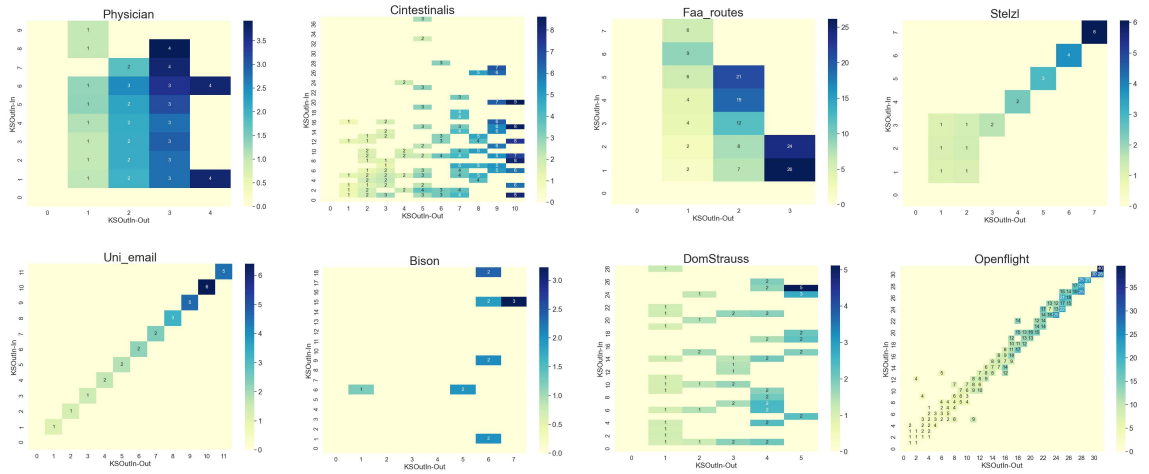


FIGURE 2. **Heatmaps showing the relationship between $KSOutIn_{In}$, $KSOutIn_{Out}$, and the average spreading capability of the corresponding shells across the eight directed networks.**

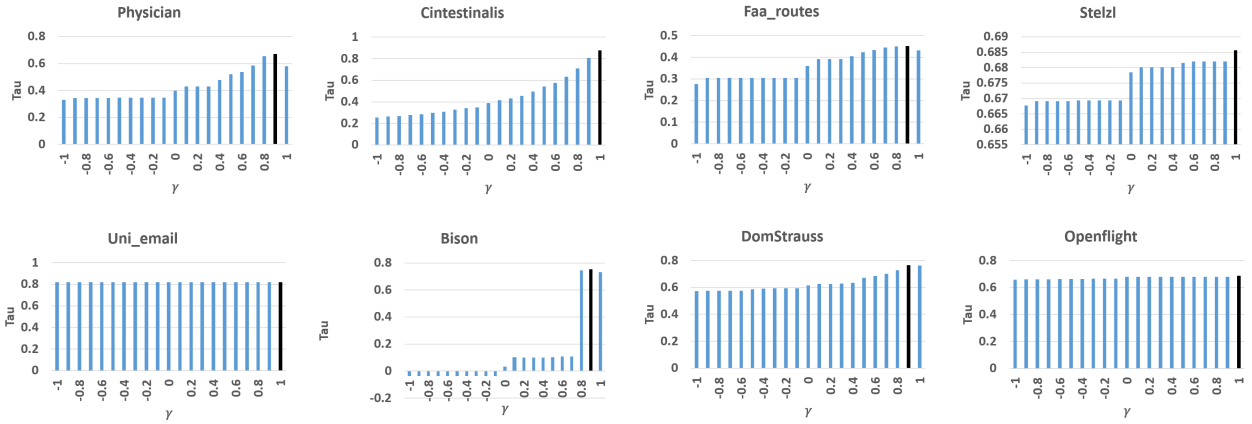


FIGURE 3. **Variation of Kendall's Tau between the ranking produced by KS_λ and the actual spreading capability for different values of λ across the eight directed networks.**

λ increases and reaches its maximum value at $\lambda = 0.9$. However, a slight decrease is observed when λ reaches 1. This behavior indicates that although $KSOutIn_{Out}$ is the dominant component in determining the ranking of nodes and shells, completely eliminating the contribution of $KSOutIn_{In}$ does not necessarily produce the strongest correspondence with spreading capability. In these networks, retaining a small contribution from $KSOutIn_{In}$ can slightly improve the resulting ranking.

In the Cintestinalis and Uni_email networks, the maximum value of Kendall's Tau is obtained at $\lambda = 1$. Thus, in these two networks, using $KSOutIn_{Out}$ alone provides the strongest correspondence with the observed spreading capability. In contrast, for the DomStrauss network, Kendall's Tau remains approximately constant at 0.821 across the entire range of λ. This result suggests that, for this network, changing the relative weights of $KSOutIn_{In}$ and $KSOutIn_{Out}$ has little to no effect on the resulting ranking.

Overall, the optimal value of λ lies in the range 0.9 to 1 for most of the networks considered. This finding is consistent with the patterns observed in the heatmaps and further supports the conclusion that $KSOutIn_{Out}$ has a stronger relationship with node spreading capability than $KSOutIn_{In}$. At the same time, the superior performance of $\lambda = 0.9$ in several networks indicates that retaining a small contribution from $KSOutIn_{In}$ can be beneficial in capturing structural differences that are not fully reflected by $KSOutIn_{Out}$ alone. Therefore, $\lambda = 0.9$ is considered as a candidate value for a transferable combined criterion that can be applied across different directed networks.

C. ADDRESSING CORE-LIKE GROUPS AND RESOLUTION LIMITATIONS

Conventional k-shell-based methods may suffer from two important limitations, namely the formation of core-like groups and limited resolution between structurally different nodes or shells. The proposed framework addresses these limitations by incorporating both the structural information provided by

KSOutIn and the propagation-related information captured by SSPI.

Among the networks considered in the dataset, the Physician network inherently exhibits the core-like group problem. To investigate the ability of SSPI to distinguish nodes within such groups, Fig. 4 illustrates the relationship between $KSOutIn_{Out}$ and SSPI for the Physician network. The horizontal axis represents $KSOutIn_{Out}$, while the vertical axis represents the SSPI value. The color of each point indicates the actual spreading capability of the corresponding node obtained using the SIR model.

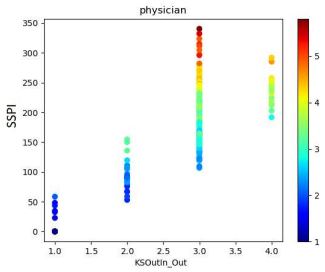


FIGURE 4. Relationship between $KSOutIn_{Out}$ and SSPI in the Physician network. The color indicates the actual spreading capability of the nodes obtained using the SIR model.

As shown in Fig. 4, nodes with similar SSPI values exhibit similar spreading capabilities, while variations in spreading capability are reflected more clearly along the $KSOutIn_{Out}$ axis. In particular, for a given SSPI value, the color distribution remains relatively consistent, indicating that SSPI provides a more discriminative characterization of nodes within the core-like groups. This observation suggests that incorporating propagation information into the KSOutIn-based structural representation can help distinguish nodes that may otherwise be assigned similar structural positions.

In addition to the core-like group problem, the resolution of the different KS-based decompositions is evaluated using the previously defined Monotonicity Index. Accordingly, higher Resolution values indicate a greater discriminative capability for ranking nodes or shells according to their spreading capability. The results are reported in Table 4. The results show that incorporating both incoming and outgoing connectivity generally improves the resolution compared with using either component individually. In particular, *KSOutIn* achieves the highest mean Resolution among the evaluated methods, with a mean value of 0.603, followed by *KSOutIn-In* with 0.596 and *KSInOut*/*KSInOut-Out* with 0.591. In comparison, *KSIn* and *KSOut* obtain mean Resolution values of 0.483 and 0.460, respectively. Overall, the results indicate that the proposed framework can improve the discrimination of nodes that exhibit similar shell-based structural characteristics. The KSOutIn decomposition provides a higher-resolution structural representation by jointly incorporating incoming and outgoing connectivity, while SSPI further incorporates propagation information to distinguish nodes within core-like groups.

D. FEATURE IMPORTANCE ANALYSIS AND MODEL EVALUATION

1) Determination of the Most Important Features

To determine the importance of the proposed features independently of the underlying model type, whether linear or nonlinear, model-agnostic interpretation methods can be employed. Several approaches have been developed for model interpretation, including Partial Dependence Plots (PDP), Local Interpretable Model-agnostic Explanations (LIME), Permutation Feature Importance (PFI), and SHAP (SHapley Additive exPlanations). Among these approaches, PFI and SHAP are widely used model-agnostic methods for evaluating feature importance.

PFI addresses the question of how much the predictive performance of a model changes when the information provided by a particular feature is disrupted. In this approach, the values of a feature are randomly permuted while the remaining features are kept unchanged, and the resulting change in model performance is measured. A larger decrease in performance indicates a greater contribution of the corresponding feature to the predictive performance of the model. Therefore, PFI provides a model-agnostic measure of the contribution of individual features.

SHAP, in contrast, is based on Shapley values from cooperative game theory and estimates the contribution of each feature to the model output by considering its contribution across different combinations of features. This provides a detailed interpretation of how individual features contribute to the predictions of a model.

In this study, both PFI and SHAP are employed to identify the most important features among the proposed KS-based features. PFI is applied to all three considered models, namely Linear Regression (LR), Random Forest (RF), and Gradient Boosting Machine (GBM), whereas SHAP is applied to the two nonlinear models, RF and GBM. The resulting feature-importance measures are used to investigate which structural characteristics contribute most strongly to the prediction of shell importance and spreading capability.

2) Model Implementation and Evaluation

Three machine-learning models, namely Linear Regression (LR), Random Forest (RF), and Gradient Boosting Machine (GBM), are implemented and evaluated using the features introduced in Section III-C. The experimental datasets consist of both real-world and synthetic networks exhibiting different structural and connectivity characteristics. The objective of this analysis is to determine whether the combination of KSOutIn-based structural descriptors, local connectivity measures, shell-level properties, and global network characteristics can improve the prediction of shell spreading capability compared with relying on a single KSOutIn measure.

To quantify the agreement between the shell ranking predicted by each model and the reference ranking determined according to spreading capability, Kendall's Tau (τ) is employed. Kendall's Tau provides a rank-based measure of association between the predicted and reference orderings, where

TABLE 4. Resolution values based on the Monotonicity Index across the benchmark networks. Higher values indicate higher resolution.

Method	Physician	Cintestinalis	Faa_routes	Stelzl	Uni_email	Bison	DomStrauss	Openflight	Mean
KSIn	0.385	0.763	0.138	0.419	0.809	0.356	0.261	0.730	**0.483**
KSOut	0.253	0.741	0.044	0.413	0.809	0.297	0.391	0.729	**0.460**
KSInOut	0.671	0.872	0.257	0.417	0.809	0.442	0.528	0.733	**0.591**
KSInOut-In	0.385	0.763	0.138	0.419	0.809	0.356	0.261	0.730	**0.483**
KSInOut-Out	0.671	0.872	0.257	0.417	0.809	0.442	0.528	0.733	**0.591**
KSOutIn	0.800	0.741	0.389	0.413	0.809	0.360	0.583	0.729	**0.603**
KSOutIn-In	0.729	0.870	0.344	0.422	0.809	0.300	0.562	0.733	**0.596**
KSOutIn-Out	0.253	0.741	0.044	0.413	0.809	0.297	0.391	0.729	**0.460**

a higher value indicates stronger agreement between the two rankings. The Kendall's Tau values obtained for the evaluated models on the test networks are reported in Tables 5–7.

To examine the generalization capability of the proposed models and, in particular, to investigate the contribution of synthetic networks to model learning, three different training and testing scenarios are considered. In Scenario I, the models are trained exclusively on synthetic networks and tested on real-world networks. Scenario II considers training and testing using real-world networks only. In Scenario III, real-world and synthetic networks are jointly used for model training, after which the resulting models are evaluated on the corresponding test networks. These three scenarios provide complementary perspectives on the extent to which the learned relationships are dependent on the type and composition of the training data.

To reduce the dependence of the evaluation on a particular train–test partition, all experiments are conducted at the network level using a Leave-One-Network-Out strategy. In each iteration, one complete network is excluded from the training set and reserved exclusively for testing. The procedure is then repeated such that every network is used once as the held-out test network. As a result, no shell from the test network is available to the model during training, allowing the evaluation to more directly assess whether the learned relationships can be transferred to previously unseen network structures.

In addition to predictive performance, the contribution of the input variables is investigated through feature-importance analysis. Two complementary measures are considered: the Agnostic Average Rank and the average linear-regression coefficient (LR Coef Avg). The corresponding results for the three training scenarios are presented in Tables 8–10. The analysis provides information not only about which variables are repeatedly identified as influential by the models, but also about the average direction and magnitude of the linear association captured by the LR model.

3) Scenario I: Training on Synthetic Networks and Testing on Real-World Networks

In the first scenario, the models are trained exclusively on synthetic networks and subsequently evaluated on real-world networks. According to Table 8, f_7*_KSOutIn*$_{Out}$ obtains the highest importance according to the Agnostic Average Rank, with an average rank of 1.37. It is followed by f_3*_meanDCOut* and f_2*_boundaryOutPerNode*, with average ranks of 2.98 and 3.87, respectively. The relatively high position of these three variables indicates that the models rely substantially on information related to the out-degree-based KS position of a shell and its outgoing connectivity. More specifically, the importance of f_3*_meanDCOut* and f_2*_boundaryOutPerNode* suggests that, beyond the shell-level KS value itself, both the average outward connectivity of shell nodes and the number of outgoing edges crossing the shell boundary contain information relevant to the spreading-based ranking.

The model-performance results reported in Table 5 further demonstrate the predictive capability of the models in this setting. The average Kendall's Tau values for LR, RF, and GBM are 0.843, 0.768, and 0.820, respectively. The obtained values indicate a generally strong agreement between the predicted shell rankings and the reference rankings across the evaluated real-world networks. For individual networks, particularly high agreement is observed in several cases. For example, in the Uni_email network, Kendall's Tau reaches 0.939 for LR and 0.970 for GBM. These results show that models trained only on synthetic networks can learn relationships between the proposed features and spreading capability that remain applicable when the models are transferred to real-world network structures. The results of Scenario I therefore provide evidence that the structural information encoded by the proposed features is not restricted to the particular synthetic networks used during training. Nevertheless, the differences in Kendall's Tau across test networks also indicate that the transferability of the learned relationships is influenced by the structural characteristics of the target network.

4) Scenario II: Training and Testing on Real-World Networks

In the second scenario, both model training and evaluation are conducted using real-world networks. As shown in Table 9, f_7*_KSOutIn*$_{Out}$ remains one of the most important variables, while f_2*_boundaryOutPerNode* occupies the next position in the feature-importance analysis. The persistence of these variables when the training data are restricted to real-world networks indicates that their contribution is not solely a consequence of the synthetic-data setting observed in Scenario I.

The predictive performance reported in Table 6 is lower on average than that obtained in Scenario I. The average Kendall's Tau values for LR, RF, and GBM are 0.778, 0.633, and 0.721, respectively. These values demonstrate that the three models remain capable of recovering a substantial por-

tion of the spreading-based ranking, although the degree of agreement varies across the test networks.

The network-level results further illustrate this variability. For instance, LR achieves a Kendall's Tau of 0.929 on the Bison network, whereas the corresponding values for the Physician and Faa_routes networks are 0.647 and 0.640, respectively. Such variation suggests that predictive performance is determined not only by the learning algorithm and the composition of the training set, but also by the structural properties of the particular network on which the model is evaluated. This observation is consistent with the use of the Leave-One-Network-Out strategy, since each test network represents a structural configuration that is completely excluded from the corresponding training iteration.

5) Scenario III: Training on Combined Real-World and Synthetic Data

In the third scenario, real-world and synthetic networks are jointly incorporated into the training data. The resulting models are then evaluated on the held-out test networks. This scenario is intended to examine whether the two data sources provide complementary structural information and whether their combination can improve the ability of the models to learn relationships that generalize across networks with different structural characteristics.

As reported in Table 10, f_7*_KSOutIn*$_{Out}$ remains among the most important variables, with an Agnostic Average Rank of 2.00. The feature f_2*_boundaryOutPerNode* follows with an average rank of 2.67. The repeated appearance of these variables near the top of the feature-importance analysis across the evaluated scenarios suggests that their contribution is comparatively stable with respect to changes in the composition of the training data. In other words, both the shell-level *KSOutIn*$_{Out}$ value and the outgoing connectivity across the shell boundary continue to provide relevant information for predicting spreading-based shell rankings.

The model-performance results presented in Table 7 show average Kendall's Tau values of 0.800, 0.735, and 0.760 for LR, RF, and GBM, respectively. Compared with Scenario II, in which only real-world networks are used for training, all three models exhibit higher average Kendall's Tau values when synthetic and real-world data are combined. This observation indicates that the additional structural patterns introduced by the synthetic networks can complement the information contained in the real-world training data. However, the magnitude and consistency of this effect vary across the individual networks, demonstrating that the benefit of combined training is dependent on the characteristics of the test network as well as on the model itself.

6) Overall Analysis of Feature Importance and Model Performance

Taken together, the results obtained under the three training scenarios provide evidence that the proposed feature set captures multiple dimensions of shell structure that are relevant to spreading-based ranking. Among the variables introduced in Section III-C, f_7*_KSOutIn*$_{Out}$ consistently appears among the highest-ranked features across the evaluated scenarios. The features f_2*_boundaryOutPerNode* and f_3*_meanDCOut* also repeatedly occupy relatively high positions in the feature-importance analysis. This pattern indicates that variables describing the out-degree-based KS position and outward directed connectivity contain substantial information for characterizing shell spreading capability.

The importance of f_7*_KSOutIn*$_{Out}$ is also conceptually consistent with its definition as the shell-level out-degree-based position in the KSOutIn decomposition. In contrast, f_2*_boundaryOutPerNode* and f_3*_meanDCOut* provide additional information that is not directly represented by the shell KS value alone. The former describes the number of outgoing connections from the shell to the remainder of the network on a per-node basis, whereas the latter captures the average outgoing degree of shell nodes. Their repeated importance therefore suggests that shell ranking may depend on both the structural position assigned by the decomposition and the extent to which the shell can directly connect to and reach other parts of the network.

The remaining features provide complementary information at different structural scales. The shell-level variables f_8*_KSOutIn*$_{In}$, f_1*_internalConn*, f_6*_diameter*, and f_5*_iterationFactor* describe, respectively, the in-degree-based KS position, internal shell connectivity, the topological extent of the shell, and its location within the node-removal process. The global variables f_9*_maxKSOutIn*$_{Out}$, f_{10}*_maxKSOutIn*$_{In}$, and f_{11}*_totalShells* provide information about the scale and structural differentiation of the complete network. The presence of these variables in the feature set allows the models to consider a shell not only as an isolated structural entity but also in relation to the broader KSOutIn decomposition of the network.

A comparison of the three training scenarios further shows that synthetic data provide useful structural information for the learning process. In the experiments considered here, training exclusively on synthetic networks yields the highest average Kendall's Tau values among the three scenarios for LR, RF, and GBM. At the same time, combining synthetic and real-world data produces higher average performance than training exclusively on real-world networks for all three models. Thus, the results do not indicate that synthetic networks merely substitute for real-world training data. Rather, they demonstrate that synthetic networks can contribute additional structural patterns, while their combination with real-world networks can improve the average predictive performance relative to a real-world-only training setting.

An important aspect of these findings is that the observed performance differences should be interpreted at the network level. The Kendall's Tau values vary considerably among the test networks even within the same scenario and for the same learning algorithm. This variation indicates that the relationships learned from the proposed features are influenced by the underlying network topology, directed connectivity, and structural organization. Consequently, a feature that is

informative on average may not contribute equally across all network types.

Finally, the use of the Leave-One-Network-Out evaluation strategy is important for interpreting the generalization results. Because the complete test network is excluded from training in each iteration, the reported Kendall's Tau values do not simply reflect the ability of the models to predict shells from networks already represented in the training data. Instead, they measure the extent to which the learned relationships between the proposed features and spreading capability can be transferred to network structures that were not observed during model fitting. The variation in performance across test networks nevertheless shows that such generalization is not independent of network structure.

Overall, the experimental results demonstrate that the proposed feature set provides a multidimensional representation of KSOutIn shells by combining direct KS information, directed connectivity, internal and topological characteristics, decomposition-stage information, and global network descriptors. Among these variables, $f_7_KSOutIn_{Out}$, $f_2_boundaryOutPerNode$, and $f_3_meanDCOut$ repeatedly emerge as influential features. At the same time, the scenario-level comparison indicates that synthetic networks can provide complementary structural information, and that their inclusion in the training process can improve predictive performance relative to real-world-only training. These findings support the use of the proposed KS-based feature representation as a basis for learning shell-level spreading capability across directed networks with heterogeneous structural characteristics.

TABLE 5. Kendall's Tau values of the LR, RF, and GBM models under Scenario I: training on synthetic networks and testing on real-world networks.

Test Network	LR_Tau	RF_Tau	GBM_Tau
Bison	**0.786**	0.429	0.714
Physician	0.822	**0.841**	0.787
Faa_routes	0.853	0.858	**0.875**
Cintestinalis	**0.867**	0.730	0.754
DomStrauss	0.817	0.818	**0.843**
Uni_email	0.939	0.879	**0.970**
Stelzl	0.829	0.842	**0.905**
Openflight	**0.832**	0.745	0.714
Average	**0.843**	0.768	0.820

TABLE 6. Kendall's Tau values of the LR, RF, and GBM models under Scenario II: training and testing on real-world networks.

Test Network	LR_Tau	RF_Tau	GBM_Tau
Bison	0.929	0.286	0.714
Physician	0.647	0.551	0.622
Faa_routes	0.640	0.780	0.794
Cintestinalis	0.837	0.597	0.602
DomStrauss	0.709	0.789	0.762
Uni_email	0.939	0.848	0.848
Stelzl	0.714	0.498	0.746
Openflight	0.808	0.718	0.681
Average	**0.778**	0.633	0.721

TABLE 7. Kendall's Tau values of the LR, RF, and GBM models under Scenario III: training on combined real-world and synthetic data.

Test Network	LR_Tau	RF_Tau	GBM_Tau
Bison	0.929	0.357	0.643
Physician	0.804	0.843	0.806
Faa_routes	0.800	0.873	0.918
Cintestinalis	0.797	0.691	0.689
DomStrauss	0.668	0.825	0.819
Uni_email	0.909	0.909	0.818
Stelzl	0.848	0.600	0.848
Openflight	0.809	0.733	0.704
LFR1	0.734	0.680	0.702
LFR2	0.769	0.814	0.816
LFR3	0.715	0.700	0.646
LFR4	0.744	0.744	0.755
LFR5	0.859	0.810	0.837
BA_m2	0.879	0.818	0.667
BA_m4	0.766	0.587	0.778
BA_m6	0.771	0.771	0.716
Average	0.800	0.735	0.760

TABLE 8. Feature-importance analysis under Scenario I based on Agnostic Average Rank and average LR coefficients.

Feature	Agnostic_Avg_Rank	LR_Coef_Avg
$f_7_KSOutIn_{Out}$	1.370	11.299
$f_3_meanDCOut$	2.981	3.588
$f_2_boundaryOutPerNode$	3.870	-0.957
$f_{11}_totalShells$	5.278	6.009
$f_8_KSOutIn_{In}$	5.704	1.476
$f_4_meanDCIn$	6.111	-1.061
$f_{10}_maxKSOutIn_{In}$	6.111	-1.754
$f_9_maxKSOutIn_{Out}$	7.074	-3.060
$f_5_iterationFactor$	7.944	0.582
$f_1_internalConn$	8.204	-0.854
$f_6_diameter$	8.889	-0.178

TABLE 9. Feature-importance analysis under Scenario II based on Agnostic Average Rank and average LR coefficients.

Feature	Agnostic_Avg_Rank	LR_Coef_Avg
$f_7_KSOutIn_{Out}$	2.354	19.782
$f_2_boundaryOutPerNode$	2.792	8.545
$f_3_meanDCOut$	4.417	8.198
$f_{11}_totalShells$	4.646	17.236
$f_{10}_maxKSOutIn_{In}$	5.833	-2.070
$f_8_KSOutIn_{In}$	5.896	5.620
$f_9_maxKSOutIn_{Out}$	6.000	-15.890
$f_4_meanDCIn$	6.188	-14.176
$f_6_diameter$	8.500	-0.399
$f_1_internalConn$	8.646	-0.208

TABLE 10. Feature-importance analysis under Scenario III based on Agnostic Average Rank and average LR coefficients.

Feature	Agnostic_Avg_Rank	LR_Coef_Avg
$f_7_KSOutIn_{Out}$	2.000	17.631
$f_2_boundaryOutPerNode$	2.667	8.410
$f_{11}_totalShells$	4.500	14.662
$f_3_meanDCOut$	5.031	6.530
$f_9_maxKSOutIn_{Out}$	5.510	-14.643
$f_4_meanDCIn$	5.792	-12.879
$f_{10}_maxKSOutIn_{In}$	6.365	-0.563
$f_8_KSOutIn_{In}$	6.396	4.523
$f_6_diameter$	8.354	0.065
$f_5_iterationFactor$	8.604	0.728

E. ACCURACY (EVALUATION OF KENDALL'S TAU)

The ranking accuracy of the proposed methods and the existing directed ranking methods was evaluated by comparing the rankings generated by these methods with the reference ranking obtained from the directed SIR model. Kendall's Tau was employed as the evaluation metric, where a higher value indicates a stronger agreement between the produced ranking and the SIR-based reference ranking. The evaluation was performed on the eight real-world directed complex networks listed in Table 1, using ten different spreading probabilities, including the threshold spreading probability. Figure 5 shows the Kendall's tau values of the evaluated ranking methods across different spreading probabilities on the eight networks. Figure 7 in the Appendix provides a zoomed-in view with a narrower y-axis range, making differences among the methods shown there easier to discern.

As shown in Figure 5, the proposed SSPI method exhibits highly competitive performance across the evaluated networks. In particular, SSPI achieves the highest Kendall's Tau over all considered spreading probabilities on the Faa_routes network. On the Openflight, Cintestinalis, Bison, and Stelzl networks, SSPI also remains highly competitive with the MAM-based methods, although the relative performance of the methods varies across the spreading-probability range.

On the Physician network, SSPI achieves the highest ranking accuracy at the threshold spreading probability and at spreading probabilities above the threshold. This result indicates that the ranking produced by SSPI is particularly consistent with the SIR-based reference ranking under these spreading conditions. In contrast, MAM_KSOutIn_DCOut achieves the highest performance on the Uni_email and DomStrauss networks, as well as on the Bison network when considering the average ranking accuracy reported in Table 1. The observed differences across networks indicate that the performance of the ranking methods depends on the structural properties of the underlying directed network and on the spreading regime.

For a more detailed comparison, the plots corresponding to Figure 5 are also provided in the appendix with an enlarged vertical range for higher Kendall's Tau values. This representation makes the differences among the competing methods more distinguishable, particularly when several methods obtain closely clustered performance values.

Table 11 reports the average Kendall's Tau values of the main competing methods over the ten considered spreading probabilities for each network. The table includes the proposed SSPI method, its reduced-feature counterpart SSPID, and the MAM variants based on *KSOutIn* and *KSInOut* combined with the corresponding degree-based features.

According to Table 11, SSPI obtains the highest average Kendall's Tau on four networks, namely Physician, Cintestinalis, Faa_routes, and Stelzl. MAM_KSOutIn_DCOut achieves the highest value on Uni_email, DomStrauss, and Bison. On Openflight, all six reported methods obtain the same average Kendall's Tau of 0.777. The aggregate results further show that SSPI obtains the largest sum, with a value of 7.006, compared with 6.984 for MAM_KSOutIn_DCOut and 6.965 for SSPID. Dividing the aggregate value by the eight networks gives an overall mean Kendall's Tau of 0.876 for SSPI. These results demonstrate that SSPI maintains a high level of agreement with the SIR-based reference ranking across different directed network structures and spreading conditions.

The comparison between SSPI and SSPID is particularly informative because SSPID corresponds to the SSPI framework with the relevant introduced feature removed. SSPI achieves a higher average Kendall's Tau than SSPID on Physician, Cintestinalis, Faa_routes, Stelzl, and in the aggregate over all eight networks. This comparison provides evidence that the introduced feature contributes positively to the ranking performance of the proposed framework.

F. PRECISION

The ranking methods were further evaluated using the Imprecision function to assess the discrepancy between the ranking produced by each method and the ranking implied by the spreading process. In contrast to Kendall's Tau, lower Imprecision values indicate better agreement with the spreading-based reference. The evaluation was conducted for 24 directed ranking methods over the eight real-world directed networks. The spreading probability p was varied from 0.01 to 0.20 with an increment of 0.01, resulting in twenty considered spreading-probability values. The corresponding results are presented in Figure 6.

Figure 6 shows that MAM_KSOutIn_DCOut generally provides very low imprecision across the evaluated networks. Among the methods reported in Table 12, it obtains the minimum imprecision on Cintestinalis, Stelzl, Uni_email, Openflight, and Bison. On Physician, SSPI achieves the lowest imprecision, while on Faa_routes the lowest reported value is obtained by MAM_KSInOut_DCIn. These results are consistent with the Kendall's Tau analysis and further illustrate that the relative performance of the methods varies across network structures and spreading probabilities. Table 12 summarizes the average Imprecision values over the considered spreading probabilities. The results in Table 12 show that SSPI has the lowest aggregate Imprecision, with a sum of 0.098 across the eight networks, followed by SSPID with 0.128 and MAM_KSOutIn_DCIn with 0.129. When the aggregate sum is normalized over the eight networks, SSPI obtains an overall mean Imprecision of approximately 0.012. This is lower than the corresponding values of MAM_KSOutIn_DCOut and SSPID, which are approximately 0.018 and 0.016, respectively.

The comparison between SSPI and SSPID is again important because SSPID excludes the corresponding introduced feature from the SSPI framework. SSPI achieves a lower average Imprecision than SSPID in the aggregate results, despite both methods obtaining identical values on several individual networks. This reduction in Imprecision indicates a closer correspondence between the SSPI ranking and the spreading-

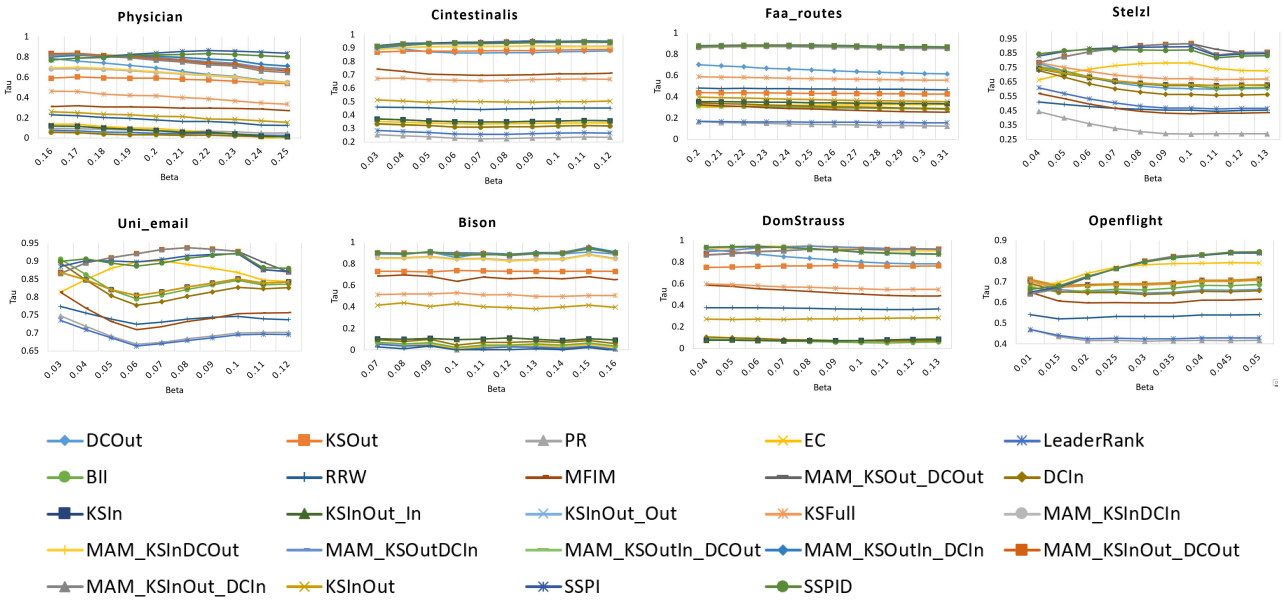


FIGURE 5. **Kendall's Tau values of the directed ranking methods for different spreading probabilities across the eight real-world directed networks.**

TABLE 11. **Average Kendall's Tau values over the considered spreading probabilities for the evaluated directed ranking methods across the eight real-world networks.**

Network	MAM_KSOutIn_DCOut	MAM_KSOutIn_DCIn	MAM_KSInOut_DCOut	MAM_KSInOut_DCIn	SSPI	SSPID	Max
Physician	0.787	0.784	0.766	0.751	0.829	0.809	0.829
Cintestinalis	0.938	0.939	0.938	0.937	0.941	0.935	0.941
Faa_routes	0.879	0.875	0.875	0.869	0.881	0.878	0.881
Stelzl	0.863	0.863	0.863	0.863	0.867	0.856	0.867
Uni_email	0.907	0.907	0.907	0.907	0.899	0.899	0.907
DomStrauss	0.925	0.924	0.905	0.904	0.911	0.911	0.925
Openflight	0.777	0.777	0.777	0.777	0.777	0.777	0.777
Bison	0.908	0.908	0.901	0.901	0.900	0.900	0.908
Sum	6.984	6.975	6.931	6.909	7.006	6.965	7.006

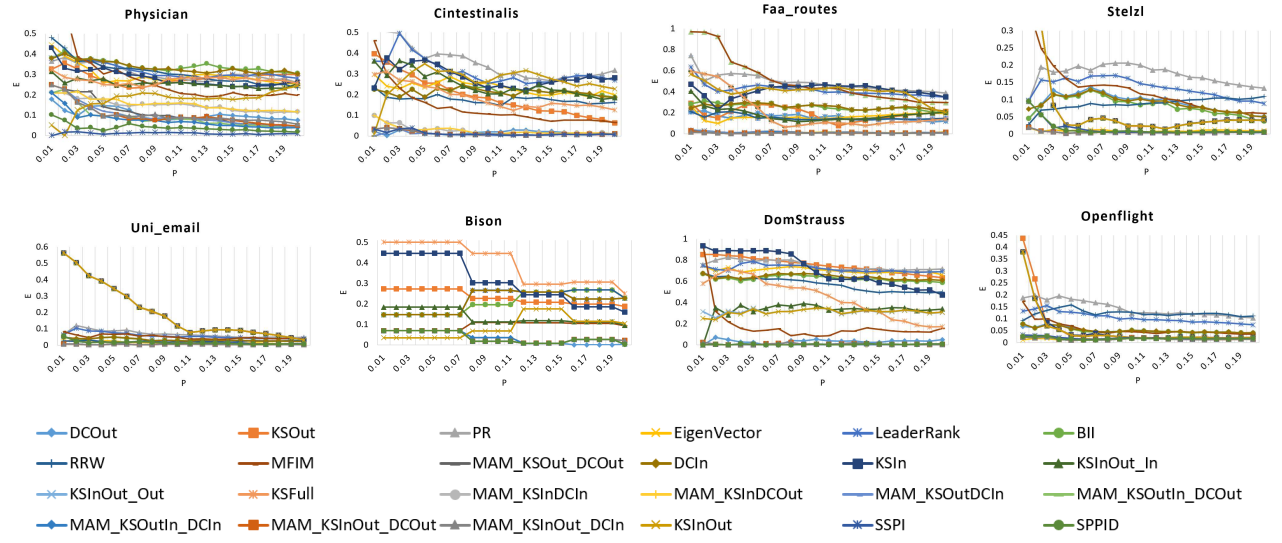


FIGURE 6. **Imprecision values of the directed ranking methods for different spreading probabilities across the eight real-world directed networks. Lower values indicate better ranking performance.**

TABLE 12. Average Imprecision values over the considered spreading probabilities for the evaluated directed ranking methods across the eight real-world networks. Lower values indicate better ranking performance.

Network	MAM_KSOutIn_DCOut	MAM_KSOutIn_DCIn	MAM_KSInOut_DCOut	MAM_KSInOut_DCIn	SSPI	SSPID	Min
Physician	0.089	0.076	0.100	0.106	0.011	0.039	0.011
Cintestinalis	0.009	0.009	0.013	0.013	0.012	0.012	0.009
Faa_routes	0.012	0.012	0.010	0.009	0.015	0.020	0.009
Stelzl	0.005	0.005	0.005	0.005	0.015	0.015	0.005
Uni_email	0.005	0.005	0.005	0.005	0.017	0.016	0.005
DomStrauss	0.002	0.003	0.008	0.009	0.006	0.006	0.002
Openflight	0.015	0.015	0.015	0.015	0.020	0.018	0.015
Bison	0.003	0.003	0.022	0.022	0.003	0.003	0.003
Sum	0.141	0.129	0.179	0.185	0.098	0.128	0.098

based reference ranking and supports the contribution of the introduced feature to the ranking framework.

Overall, the Kendall's Tau and Imprecision analyses provide complementary evidence regarding the behavior of the evaluated methods. The Kendall's Tau results assess the agreement between the produced and reference rankings directly, whereas the Imprecision analysis examines the deviation of the ranking from the spreading-based performance. Across the eight evaluated directed networks, SSPI achieves the highest aggregate Kendall's Tau and the lowest aggregate Imprecision among the methods reported in Tables 11 and 12. The consistent improvement of SSPI over SSPID in the aggregate results further highlights the contribution of the introduced feature to the proposed ranking framework.

G. COMPUTATIONAL COMPLEXITY

The computational complexity of the proposed framework and the benchmark methods is summarized in Table 13. For the proposed KSOutIn decomposition, the computation is performed using adjacency-list representations, incremental degree updates, and bucket-based processing. Each node is processed a constant number of times, while each directed edge is examined only when its removal-induced degree update is required. Hence, the total cost of node and edge processing is $O(n)$ and $O(m)$, respectively, resulting in an overall complexity of $O(n+m)$. For sparse networks, where $m = O(n)$, this becomes $O(n)$, demonstrating the linear scalability of the proposed decomposition with respect to network size. The same implementation principle applies to the KSInOut decomposition, while the additional operations required by the MAM- and SSPI-based methods introduce a dependence on the average degree $\langle k \rangle$, as reflected in Table 13.

For iterative methods such as PageRank and LeaderRank, the cost of each iteration is proportional to the number of edges, leading to $O(m*Iter)$ overall complexity when the initialization cost $O(n)$ is dominated by iterative edge processing. The quadratic methods, including BII and RRW, require $O(n^2)$ operations, while MFIM and the MAM-based methods additionally depend on the average degree $\langle k \rangle$. Accordingly, Table 13 provides a unified comparison of the computational requirements of all considered methods.

VI. CONCLUSION AND FUTURE WORK

This study presented a feature-guided framework for influential-node identification in directed complex networks. The proposed *KSOutIn* representation preserves both incoming and outgoing shell information and provides a joint description of directional coreness. The analysis showed that the two shell components alone are not sufficient to establish a reliable ordering of directed shells, motivating the use of additional structural information. Through learning-based feature analysis, the proposed framework identifies several informative characteristics, particularly the out-oriented shell position, outgoing connectivity, and shell-level structural properties. These features are incorporated into the proposed *SSPI* ranking mechanism to improve the discrimination of nodes with similar shell positions.

Experiments on eight real-world directed networks demonstrate that SSPI achieves an overall mean Kendall's Tau of 0.876 and a mean Imprecision of approximately 0.012. In addition, KSOutIn achieves a mean Monotonicity of 0.603, indicating improved ranking resolution compared with single-direction shell representations. The proposed shell decomposition also has $O(n+m)$ computational complexity, supporting its applicability to sparse directed networks.

Future research can investigate adaptive feature selection and propagation mechanisms that reduce the dependence on manually defined structural descriptors and parameters. The framework can also be extended to weighted, temporal, multilayer, and signed directed networks. Finally, evaluating the proposed method on larger and more heterogeneous real-world networks and under alternative spreading processes would provide further evidence regarding its generalization and robustness.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for valuable feedback that improved this manuscript.

APPENDIX

A. PRELIMINARIES

This section briefly introduces the methods used in this study, including the directed SIR model, K-shell decomposition, learning models, and model-agnostic feature interpretation methods.

TABLE 13. Computational complexity of the considered methods for directed networks. Here, *n* and *m* denote the number of nodes and directed edges, respectively, $\langle k \rangle$ denotes the average degree of the network, and *Iter* represents the number of iterations required by an iterative method.

Method(s)	Time Complexity
DCIn, DCOut	$O(m)$
KSIn, KSOut, KSFull	$O(m)$
KSInOut, KSOutIn	$O(n+m)$
MAM_KSInDCIn, MAM_KSInDCOut, MAM_KSOutDCIn, MAM_KSOutDCOut	$O\left(m\langle k\rangle^2\right)$
MAM_KSOutIn_DCOut, MAM_KSOutIn_DCIn, MAM_KSInOut_DCOut, MAM_KSInOut_DCIn	$O\left((n+m)\langle k\rangle^2\right)$
SSPI, SSPID	$O\left((n+m)\langle k\rangle^2\right)$
PageRank (PR)	$O(m * Iter)$
LeaderRank	$O(m * Iter)$
Eigen vector Centrality (EC)	$O(m \log n)$
BII, RRW	$O(n^2)$
MFIM	$O\left(n\langle k\rangle^2\right)$

1) Directed SIR Model

The Susceptible–Infected–Recovered (SIR) model is used to evaluate the spreading ability of nodes in directed networks [12]. An infected node transmits the infection to each susceptible out-neighbor with probability β. The spreading threshold is given by

$$\beta_{th} = \frac{\langle DC_{\text{in}} \rangle}{\langle DC_{\text{Out}} DC_{\text{in}} \rangle}, \tag{22}$$

where DC_{in} and DC_{Out} denote the in-degree and out-degree, respectively. The spreading probability is selected such that $\beta > \beta_{th}$ ensuring a large-scale spreading regime. For each initially infected node, the SIR process is repeated independently using Monte Carlo simulations. The average number of recovered nodes over these realizations is used as the node's spreading ability.

2) K-Shell Decomposition Method

The K-shell decomposition method is an iterative procedure for decomposing a network into concentric layers. Initially, all nodes whose degree is less than or equal to one are removed from the network and assigned a k-shell index of one. Removing these nodes may reduce the degree of their neighboring nodes. If the degree of a remaining node becomes less than or equal to one, it is also removed and assigned to the same shell.

Once this process reaches a stable state, the shell index is increased by one, and the same procedure is repeated for the remaining nodes. More generally, at shell index k, all nodes whose current degree is less than or equal to k are iteratively removed until no such nodes remain. The decomposition continues until all nodes have been assigned to a shell.

Consequently, each node is assigned a k-shell index according to the shell in which it is removed. Nodes with larger k-shell indices are located in the inner layers of the network and are generally regarded as belonging to its structural core.

Several directed variants of the K-shell decomposition method have been introduced in [14]. These variants explicitly account for the directionality of edges. In particular, *KSOut* considers outgoing edges when performing the decomposition, *KSIn* considers incoming edges, and *KSInOut* simultaneously incorporates both incoming and outgoing edges.

3) Linear and Nonlinear Learning Models

Machine learning models can be used to investigate the relationship between network features and node spreading ability. Linear models, such as linear regression (LR) [18], are simple and interpretable but have limited ability to capture nonlinear relationships. Nonlinear models, such as random forest (RF) [19] and gradient boosting (GBM) [20], can capture complex patterns and feature interactions but are less directly interpretable.

In this study, LR, RF, and GBM are employed as complementary models. LR provides a directly interpretable baseline, while RF and GBM capture nonlinear relationships. SHAP and permutation feature importance (PFI) are further used to interpret feature contributions independently of model structure.

4) Linear Regression (LR)

Linear regression [18] models a continuous target as a linear combination of the input features:

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_j x_j, \tag{23}$$

where β_0 is the intercept, x_j is the jth feature, and β_j is its coefficient. The coefficients provide a direct indication of the relationship between features and the target. However, LR cannot adequately capture complex nonlinear relationships and feature interactions.

5) Random Forest (RF)

Random Forest [19] is an ensemble of decision trees constructed using bootstrap samples and randomly selected feature subsets. For regression, the final prediction is obtained by averaging the predictions of the individual trees. RF can capture nonlinear relationships and feature interactions while reducing the overfitting risk of individual trees. Its feature importance can be examined using methods such as PFI and SHAP.

6) Gradient Boosting (GBM)

Gradient Boosting constructs decision trees sequentially, with each new tree focusing on reducing the errors of the existing ensemble [20]. This approach can capture complex nonlinear relationships and interactions. Its performance depends on hyperparameters such as the number of trees, tree depth, and learning rate. In this study, GBM complements LR and RF by providing an additional nonlinear modeling framework.

B. MODEL-AGNOSTIC INTERPRETATION METHODS

Model-agnostic methods explain feature contributions without relying on the internal structure of a specific learning model. Here, SHAP and permutation feature importance (PFI) are used as complementary interpretation methods.

1) SHAP

SHAP (SHapley Additive exPlanations) [22] quantifies the contribution of each feature to a prediction relative to a baseline using Shapley values. A SHAP value indicates both the magnitude and direction of a feature's contribution. SHAP can provide local explanations for individual predictions and global interpretations by aggregating feature contributions across observations.

2) Permutation Feature Importance (PFI)

PFI [21] measures feature importance by randomly permuting one feature while keeping the others unchanged and evaluating the resulting change in model performance. A larger performance deterioration indicates greater feature importance. Unlike SHAP, PFI primarily provides a global measure of feature importance. In this study, PFI and SHAP are used together to identify features that consistently contribute to predicting node spreading ability.

REFERENCES

[1] Y. He, Y. Jin, J. Cao, S. Sui, J. Wang, and B. Ran, "Identification of key nodes in urban bus-metro network: A nk-shell algorithm based neighborhood ks," *Reliability Engineering & System Safety*, vol. 264, p. 111448, 2025.

[2] S. Esfandiari, "Enhancing data quality by identifying influential nodes: Integrating complementary features with cocofiso," in *Companion Proceedings of the ACM on Web Conference 2025*, 2025, pp. 2116–2119.

[3] X. Su, P. Hu, D. Li, B. Zhao, Z. Niu, T. Herget, P. S. Yu, and L. Hu, "Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning," *Nature biomedical engineering*, vol. 9, no. 3, pp. 371–389, 2025.

[4] S. Esfandiari and S. M. Fakhrahmad, "Mining influential spreaders in complex networks by an effective combination of the degree and k-shell," in *2024 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, 2024.

[5] S. Esfandiari and S. M. Fakhrahmad, "Uncertainty-aware fuzzy centrality measures for influential node identification: A structural modeling approach toward e-commerce applications," in *2026 12th International Conference on Web Research (ICWR)*. IEEE, 2026, pp. 173–178.

[6] S. Esfandiari and S. M. Fakhrahmad, "Identifying influential users in social networks with an extended hybrid global structure model," in *2024 10th International Conference on Web Research (ICWR)*. IEEE, 2024, pp. 1–5.

[7] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[8] Y. Ruan, X. Qin, S. Liu, M. Zhang, J. Tang, Y. Guo, and T. Yu, "Grain: Gravity-resistance adaptive framework for identifying influential nodes using multi-order structural diversity," *Information Processing & Management*, vol. 63, no. 4, p. 104618, 2026.

[9] S. Esfandiari and S. M. Fakhrahmad, "The collaborative role of k-shell and pagerank for identifying influential nodes in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 658, p. 130256, 2025.

[10] D. Chen, L. Lü, M.-S. Shang, Y.-C. Zhang, and T. Zhou, "Identifying influential nodes in complex networks," *Physica a: Statistical mechanics and its applications*, vol. 391, no. 4, pp. 1777–1787, 2012.

[11] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," Stanford Digital Library Technologies Project, Tech. Rep., 1998.

[12] Y.-L. Lee, Y.-F. Wen, W.-B. Xie, L. Pan, Y. Du, and T. Zhou, "Identifying influential nodes on directed networks," *Information Sciences*, vol. 677, p. 120945, 2024.

[13] S. Esfandiari, "A multi-attribute model to identify influential nodes in complex networks: Addressing the k-shell limitations," in *Companion Proceedings of the ACM on Web Conference 2025*, 2025, pp. 3092–3095.

[14] S. Esfandiari and M. Fakhrahmad, "A novel k-shell-based feature for detecting influential nodes in directed complex networks: Addressing resolution and core-like groups problems," *Information Processing & Management*, vol. 63, no. 6, p. 104719, 2026.

[15] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[16] S. Esfandiari and S. M. Fakhrahmad, "Predicting node influence in complex networks by the k-shell entropy and degree centrality," in *Companion Proceedings of the ACM on Web Conference 2024*, 2024, pp. 629–632.

[17] S. Esfandiari and M. R. Moosavi, "Identifying influential nodes in complex networks through the k-shell index and neighborhood information," *Journal of Computational Science*, vol. 84, p. 102473, 2025.

[18] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, "Linear regression," in *An introduction to statistical learning: With applications in python*. Springer, 2023, pp. 69–134.

[19] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004.

[20] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[21] A. Altmann, L. Tolo i, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.

[22] J. T. Hancock, T. M. Khoshgoftaar, and Q. Liang, "A problem-agnostic approach to feature selection and analysis using shap," *Journal of Big Data*, vol. 12, no. 1, p. 12, 2025.

[23] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[24] P. Bonacich, "Power and centrality: A family of measures," *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[25] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," *PloS one*, vol. 6, no. 6, p. e21202, 2011.

[26] M. Curado, L. Tortosa, and J.-F. Vicent, "A novel measure to identify influential nodes: Return random walk gravity centrality," *Information Sciences*, vol. 628, pp. 177–195, 2023.

[27] Y. Wang, L. Zhang, J. Yang, M. Yan, and H. Li, "Multi-factor information matrix: A directed weighted method to identify influential nodes in social networks," *Chaos, Solitons & Fractals*, vol. 180, p. 114485, 2024.

[28] M. F. Mukhtar, Z. Abal Abas, A. S. Baharuddin, M. N. Norizan, W. F. W. Wan Fakhruddin, W. Minato, A. H. Abdul Rasib, Z. Z. Abidin, A. F. N. Abdul Rahman, and S. H. Hairol Anuar, "Integrating local and global information to identify influential nodes in complex networks," *Scientific Reports*, vol. 13, p. 11411, 2023.

[29] S. Esfandiari and S. M. Fakhrahmad, "Identifying influential nodes in complex networks by adjusted feature contributions and neighborhood impact," *The Journal of Supercomputing*, vol. 81, no. 3, 2025.

[30] C. Giatsidis, D. M. Thilikos, and M. Vazirgiannis, "D-cores: measuring collaboration of directed graphs based on degeneracy," *Knowledge and information systems*, vol. 35, no. 2, pp. 311–343, 2013.

[31] X. Liao, Q. Liu, J. Jiang, X. Huang, J. Xu, and B. Choi, "Distributed d-core decomposition over large directed graphs," *arXiv preprint arXiv:2202.05990*, 2022.

[32] A. L. Brophy, "An algorithm and program for calculation of kendall's rank correlation coefficient," *Behavior Research Methods, Instruments, & Computers*, vol. 18, no. 1, pp. 45–46, 1986.

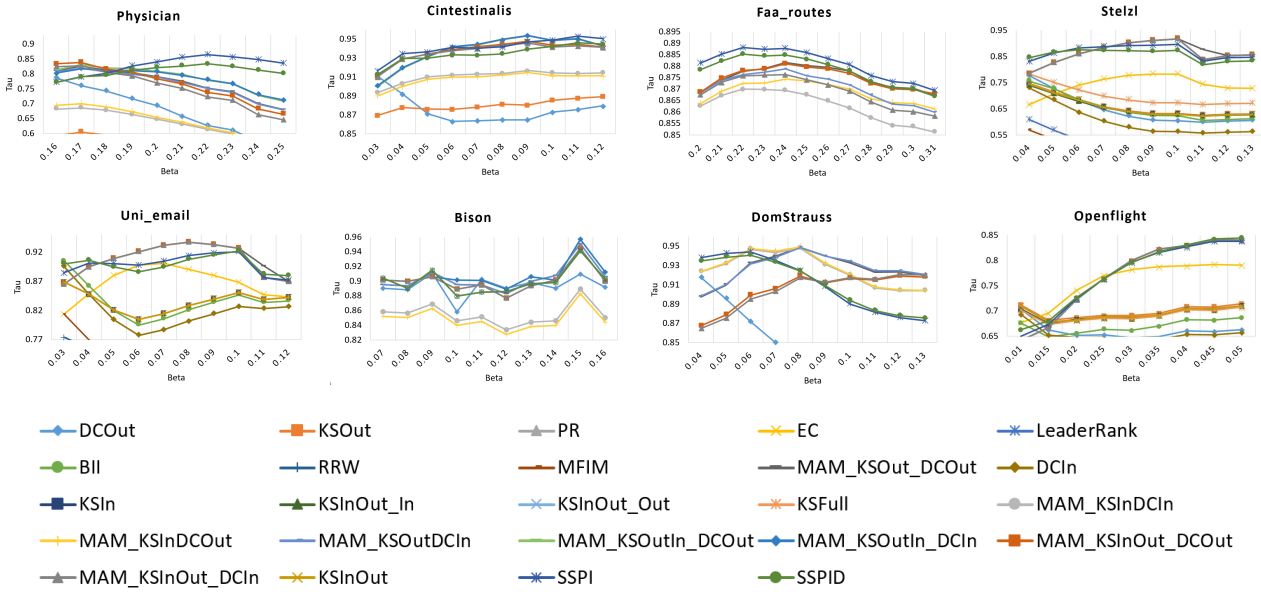


FIGURE 7. **Kendall's Tau values of the directed ranking methods for different spreading probabilities across the eight real-world directed networks.**

[33] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[34] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical Review E*, vol. 78, no. 4, p. 046110, 2008.
[35] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, vol. 30, no. 1/2, pp. 81–93, 1938.
[36] J. Bae and S. Kim, "Identifying and ranking influential spreaders in complex networks by neighborhood coreness," *Physica A: Statistical Mechanics and its Applications*, vol. 395, pp. 549–559, 2014.