

Predicting Centrality Measures in Complex Networks Utilizing Markov Chains

Anchal Gera*, Jens Dörpinghaus^{†‡§¶}, Robert Rockenfeller*

* Mathematical Institute, University of Koblenz, Germany

[†] Department of Computer Science, University of Koblenz, Germany

[‡] Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

[§] Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden

[¶] University of Pretoria, South Africa

Email: anchalgera@uni-koblenz.de, jens.doerpinghaus@bibb.de, rrockenfeller@uni-koblenz.de

Abstract—Centrality measures are of critical importance for the identification of influential nodes in networks. However, classical metrics such as degree and betweenness are static and thus fail to capture temporal dynamics. The present paper introduces a probabilistic framework for modeling centrality as a stochastic process using Markov chains. By discretizing centrality values into states, transition probability matrices are learned from temporal data, enabling future centrality prediction. First-order and higher-order models were evaluated, with the results indicating that higher-order models improve accuracy, particularly for betweenness. The results of the study indicate that centrality is inherently dynamic and probabilistic. The proposed framework offers an interpretable method to forecast node importance and distinguish stable from volatile influencers in evolving networks.

I. INTRODUCTION

NETWORKS are foundational components of complex systems in nature, technology, and society, influencing the dissemination of information, the propagation of diseases, and the distribution of resources. Network science is a field of study that utilizes centrality measures to analyze such systems, with these measures quantifying the importance of nodes within a network. It is evident that classical metrics, incorporating degree, betweenness, closeness, and eigenvector centrality, have yielded valuable insights into static network structures. However, in actual practice, networks are seldom static, for example social networks [1], [2], [3], [4], biomedical networks [5], [6], [7] or education networks [8]. Phenomena evolve, adapt and fluctuate over time, thus rendering traditional centrality measures insufficient for capturing the temporal and probabilistic nature of influence and interaction [9].

Observations of natural and engineered systems suggest that complex structures often evolve according to underlying patterns governed by physical laws or mathematical principles, often formulated in the form of differential equations [10]. Concurrently, a substantial proportion of these systems can be naturally represented as dynamic networks, where interactions between components undergo alteration over time. Examples of such phenomena include ant colonies forming trails from initially random movements, leaf venation emerging through local growth rules, and social systems organizing through repeated interactions [11], [12]. When observed through the

lens of a network, such systems demonstrate a tendency towards self-organization, albeit constrained by parameters that operate across specific temporal and spatial scales. Thus, the central question guiding this study is whether temporal measurements of network structure, governed by simple probabilistic rules, can provide a comprehensive explanation of interaction patterns and enable the prediction of future network evolution.

A significant gap in extant research is the limited ability to predict how network structures and node importance evolve over time. While a substantial body of research has focused on describing network snapshots, fewer studies have addressed the prediction of centrality dynamics in evolving networks, see [9] and the literature review. A key issue is the absence of flexible frameworks that integrate probabilistic modeling with centrality analysis to quantify influence and volatility in dynamic settings.

The present study addresses this gap by investigating centrality from a probabilistic perspective. Recent advances in the field of network analysis have seen the integration of Markov chains, and it is this foundation that the present study builds upon. The investigation of stochastic transition models as a means of capturing temporal dependencies in network evolution is the primary focus of this study.

II. LITERATURE REVIEW

The study of complex networks has evolved into a central paradigm for understanding systems across disciplines, including sociology, biology, and computer science. Foundational concepts and terminology in network science are comprehensively introduced in the work of [13], which emphasizes networks as a universal framework for representing interacting entities. Historically, the field developed through interdisciplinary contributions, ranging from early sociological theories to formal mathematical modeling and computational analysis [14].

A core problem in network science is the identification of important or influential nodes. This has led to the development of centrality measures, which quantify node importance from different structural perspectives. Classical centrality measures

include degree, betweenness, closeness, and eigenvector centrality, each capturing a distinct notion of influence [15]. Degree centrality, one of the earliest and simplest measures, reflects the number of direct connections of a node and was formalized by [16]. Betweenness centrality captures the extent to which a node lies on shortest paths between other nodes, thereby measuring its role as an intermediary [17]. Closeness centrality quantifies how close a node is to all others in terms of shortest path distances, while eigenvector centrality assigns importance based on connections to other highly connected nodes [15].

While these measures have been widely applied, they are fundamentally limited to static network representations. Real-world systems, however, are inherently dynamic, with network structures evolving over time. Recent work highlights that applying static centrality measures to evolving networks can lead to significant inaccuracies and misinterpretations [9]. This has motivated the development of temporal network analysis, where networks are represented as sequences of time-dependent snapshots.

Early approaches to dynamic networks extended static centrality measures by computing them independently at each time step. Although straightforward, this approach neglects temporal dependencies and fails to capture the continuity of node importance. More advanced methods incorporate time explicitly into centrality definitions. For instance, dynamic PageRank models treat centrality as a time-dependent dynamical system driven by evolving transition matrices [18]. These approaches enable faster detection of emerging influential nodes by emphasizing recent structural changes.

Parallel to temporal extensions, probabilistic approaches have gained importance in network analysis. Probabilistic graphical models, such as Bayesian networks, provide a principled framework for representing uncertainty and conditional dependencies between nodes [19]. These models allow for robust inference in the presence of noisy or incomplete data and enable the identification of statistically significant network structures.

Dynamic Probabilistic Networks (DPNs) extend these ideas to temporal settings by modeling network evolution as a stochastic process under the Markov assumption [20]. In this framework, the state of the network at time $t + 1$ depends only on its state at time t , allowing for tractable learning and inference. Such models have been successfully applied in domains such as biological network analysis and social interaction modeling.

Markov chains provide a particularly powerful tool for modeling stochastic processes on networks. In this setting, transitions between nodes are governed by a probability matrix, and long-term behavior is characterized by a stationary distribution. This perspective forms the basis of PageRank and related centrality measures. Research has shown that in certain network structures, particularly scale-free networks, some nodes exhibit remarkable stability in their rankings despite structural perturbations [21].

Recent work has extended Markov-based models to cap-

ture temporal dynamics and improve predictive performance. Temporal PageRank incorporates time-dependent transition matrices, enabling the modeling of evolving influence patterns [18]. Furthermore, higher-order Markov models have been proposed to account for temporal dependencies beyond the memoryless assumption, demonstrating improved accuracy in predicting centrality trajectories in dynamic networks [22].

Notwithstanding these advances, substantial gaps persist within the relevant literature. The majority of existing approaches concentrate on describing centrality at specific time points rather than predicting its evolution. Furthermore, there is a lack of integration between probabilistic modeling and centrality analysis within a unified framework. Finally, centrality is frequently regarded as a deterministic quantity, despite its inherently stochastic nature in dynamic environments.

The present paper builds upon these lines of research by introducing a probabilistic framework for centrality analysis in evolving networks. The objective of this study is to integrate the principles of Markov chain modeling with the utilization of discretized centrality trajectories. This approach is intended to facilitate the capture of both the temporal dynamics and the associated uncertainty of node importance. Furthermore, a flexible weighted centrality measure is proposed, allowing customization based on application-specific priorities. This enables the identification of both stable and volatile nodes. This approach contributes to a deeper understanding of how influence emerges and evolves in complex, dynamic systems.

III. METHODOLOGICAL BACKGROUND

The incorporation of probability theory into complex network analysis has been shown to have a significant impact on our capacity to model uncertainty, characterize random structural evolution, and quantify dynamic interdependence among nodes on a large scale, see [19], [20]. Whilst classical graph theory is typically predicated on deterministic assumptions, probabilistic frameworks explicitly acknowledge the stochastic nature of real-world networks. This paradigm shift enables researchers to move beyond static representations and address three central challenges:

- 1) Modeling incomplete or noisy observational data through uncertainty quantification,
- 2) Capturing temporal restructuring mechanisms in evolving networks,
- 3) Understanding how local stochastic interaction rules give rise to emergent global patterns.

At the core of probabilistic network modeling lies the conceptualization of networks as probabilistic graphical models. Bayesian networks provide a principled framework for representing conditional dependencies between nodes, while avoiding overfitting through likelihood-based model selection [19]. In the context of a network comprising random variables $X_1, \dots, X_n$, the joint probability distribution can be factorized as follows:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{Pa}(X_i)),$$

where $\text{Pa}(X_i)$ denotes the parent set of node X_i . This factorization induces sparse dependency structures that exclude redundant correlations and improve interpretability [19].

Probabilistic formulations are particularly powerful for dynamic networks through *Dynamic Probabilistic Networks* (DPNs) [20]. These extend the Markov assumption to the temporal domain:

$$P(X[t+1] \mid X[0], \dots, X[t]) = P(X[t+1] \mid X[t]),$$

where temporal transitions are governed by a time-invariant dependency structure. Learning such models from data involves maximizing a Bayesian score

$$\text{Score}(G) = \log P(D \mid G) + \log P(G),$$

where D represents observed temporal snapshots and $P(G)$ encodes prior structural knowledge [20]. This approach enables the joint inference of static dependencies and dynamic transition rules, with applications ranging from biological pathway modeling to social contagion analysis.

Centrality measures undergo a fundamental reinterpretation under probabilistic network models. Consider Katz centrality,

$$C_i = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji},$$

where α is a damping factor controlling the contribution of longer paths [23]. In random networks, C_i becomes a random variable whose distribution depends on the degree distribution $p(k)$. For sparse, locally tree-like networks, cavity methods yield the approximation

$$P_C(c) \approx \sum_{k=0}^{\infty} p(k) \delta\left(c - \frac{1}{1 - \alpha k \mathbb{E}[C]}\right),$$

revealing a multimodal structure in which each mode corresponds to a distinct degree class [23]. This result explains empirically observed correlations between degree and centrality, as nodes of different degrees occupy distinct probabilistic centrality regimes.

Probabilistic network theory therefore provides a powerful analytical lens, enabling robust uncertainty quantification, dynamic modeling, and improved interpretability across domains ranging from protein interaction networks to resilient infrastructure design. This theoretical framework provides us with one more idea that is "Influence is uncertain and evolves probabilistically".

A. Markov Chains and Transition Matrices in Network Centrality

Markov chain theory offers a mathematically rigorous framework for modeling stochastic transitions and predicting node importance in evolving networks. In this setting, network navigation is represented as a memoryless random walk governed by a transition matrix P , where P_{uv} denotes the probability of transitioning from node u to node v in one

step [24]. For a network with adjacency matrix A , a common construction is

$$P_{uv} = \frac{A_{uv}}{k_u},$$

where k_u is the out-degree of node u , ensuring row-stochasticity $\sum_v P_{uv} = 1$ [24], [21].

The stationary distribution π , satisfying $\pi = \pi P$, forms the theoretical basis of PageRank centrality. Introducing a teleportation parameter α yields

$$\pi = (1 - \alpha)v + \alpha\pi P,$$

where v is the teleportation distribution [18]. In scale-free networks, this formulation gives rise to *super-stable* nodes whose rankings remain robust under topological perturbations due to their disproportionate influence on π [21]. A complementary global connectivity metric is the Kemeny constant [25],

$$K = \sum_{j=2}^n \frac{1}{1 - \lambda_j},$$

where λ_j are the non-principal eigenvalues of P . Lower values of K correspond to faster mixing and more efficient information diffusion [24].

In dynamic networks, the transition matrix becomes time-dependent, leading to non-homogeneous Markov chains [18]. Temporal PageRank models this evolution via the dynamical system

$$x'(t) = (1 - \alpha)v(t) - (I - \alpha P(t))x(t),$$

where $x(t)$ represents instantaneous centrality. Empirical evaluation on Wikipedia traffic data shows that this approach detects emerging trends substantially faster than static PageRank by emphasizing recent transitions [18].

Prediction of future centrality values can be achieved using higher-order Markov chains that condition transitions on multi-step histories [22]. For a centrality time series C_t , a K -order model estimates

$$P(C_{t+1} \mid C_t, \dots, C_{t-K}) = \prod_{k=0}^K P(C_{t+1} \mid C_{t-k}),$$

with parameters learned via maximum likelihood estimation. Applied to mobile social networks, this approach achieves high prediction accuracy by capturing periodic interaction patterns, see [22].

Recent work integrates Markov transitions with graph neural networks by learning transition matrices P_θ parameterized by node features [26]. The Transition Probability Matrix (TPM) method computes node representations as

$$\phi(v) = \text{MLP}\left(\sum_{w \in N(v)} P_\theta(v, w) h_w\right),$$

where h_w denotes node embeddings. This hybrid approach demonstrates improved performance over classical Graph Convolutional Networks (GCNs) in cross-network

classification tasks and Here, MLP denotes a multi-layer perceptron that applies a nonlinear transformation to the Markov-weighted neighborhood aggregation, enabling expressive node representations beyond linear diffusion [26].

Finally, dynamic-sensitive centrality analysis shows that temporal correlation between consecutive snapshots governs prediction error bounds [27]:

$$\epsilon \leq \frac{1-\rho}{\sqrt{T}} \sum_{t=1}^T \|P_t - P_{t-1}\|_F,$$

where ρ is the temporal correlation coefficient and T is the observation window. High correlation ensures robustness even under substantial missing data, highlighting the suitability of Markov-based approaches for real-world temporal networks such as epidemic contact graphs and IoT systems.

B. Upper and lower Bounds for Centrality Measures

Several upper and lower bounds for centrality measures have been proposed in literature, see [28], [29]. Here, the authors defined $\mathfrak{G}^p = \{G_1, \dots, G_\ell\}$ as a series of graphs and $p \in \mathbb{R}$ with $0 \leq p \leq 1$ and

$$|(V(G_i) \cup V(G_{i+1})) \setminus (V(G_i) \cap V(G_{i+1}))| \leq p|V(G_i)|,$$

$$|(E(G_i) \cup E(G_{i+1})) \setminus (E(G_i) \cap E(G_{i+1}))| \leq p|E(G_i)|,$$

for $i \in \{1, \dots, \ell-1\}$. As $\mathfrak{G}^p$ now is a series of graphs with a fixed set of differences and changes from one to the other, the authors proposed approximations of the changes over time, or the error in the centrality measures that can occur due to these changes. Unless otherwise noted, we will consider $\mathfrak{G}^p = \{G_1, \dots, G_\ell\}$.

Theorem III.1. *Let $i \in \{1, \dots, \ell-1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then it holds that*

$$dc^{i+1}(v) \geq \frac{d^i(v) - p|V(G_i)|}{|V(G_i)| - 1 + p|V(G_i)|},$$

$$dc^{i+1}(v) \leq \frac{d^i(v) + p|V(G_i)|}{|V(G_i)| - 1 - p|V(G_i)|}.$$

For betweenness centrality, we define

$$\sigma = |N(G_i)|p,$$

$$\epsilon = \min\{D(G_i)^2, 2|V(G_i)|p\},$$

where $D(G)$ is the diameter of G . [28] provided two lemmata to obtain a bound for $bc^{i+1}(v)$ for $v \in V$.

Lemma III.2. *Let $i \in \{1, \dots, \ell-1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$P_v(k, j) \frac{1}{\sigma} \leq P_v^{i+1}(k, j)$$

holds.

Lemma III.3. *Let $i \in \{1, \dots, \ell-1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$P_v^{i+1}(k, j) \leq \begin{cases} P_v(k, j)\epsilon & P_v(k, j) > 0 \\ D(G_i)^2\epsilon & P_v(k, j) = 0 \end{cases}$$

holds.

Theorem III.4. *Let $i \in \{1, \dots, \ell-1\}$ so that $v \in V(G_i)$ and $v \in V(G_{i+1})$. Then,*

$$bc^i(v)\epsilon \leq bc^{i+1}(v) \leq bc^i(v) \frac{1}{\sigma}$$

holds.

For details and proofs we refer to [28], [29]. Clearly, this notation is equivalent to $\{G_t = (V_t, E_t)\}_{t=1}^T$ as a sequence of network snapshots over T discrete time steps, see Section IV-B.

IV. METHOD

A. Generation of Networks

All network simulations are implemented in Python using the *NetworkX* library, which provides data structures and algorithms for constructing, manipulating, and analyzing complex networks [13]. In order to systematically study the effects of topology on centrality dynamics, a range of canonical network models must first be generated. These models will serve as controlled environments for the analyses to be conducted.

Random networks are constructed using the Erdős-Rényi model $G(n, p)$. In this model, each of the nodes, n in number, is connected independently with probability p . This process yields homogeneous degree distributions characterized by minimal clustering and the absence of intrinsic hubs. Consequently, a neutral baseline is established, providing a reference point for the subsequent comparison of more structured topologies. Scale-free networks are generated via the Barabási-Albert preferential attachment mechanism, whereby new nodes attach to existing nodes with probability proportional to their degree. This phenomenon gives rise to heterogeneous topologies comprising a small number of high-degree hubs, which mirrors numerous real-world systems, including social and citation networks. Small-world networks are generated using the Watts-Strogatz model, which begins with a regular lattice and randomly rewires edges with probability p_{rewire} . This process produces high local clustering while maintaining short average path lengths. Such networks have been shown to capture structural features that are relevant to social and biological systems, where influence can rapidly propagate within and across communities.

To model temporal evolution, we construct dynamic networks as sequences of snapshots, denoted by $G_t = (V_t, E_t)$. Each simulation commences with a modestly connected seed graph, which is generally initiated using an Erdős-Rényi topology. At each discrete time step, a new node is added and connected to a predefined number of existing nodes, sampled from a distribution. Concurrently, a limited number of edges can undergo removal or reconfiguration to emulate network

TABLE I: Experiment 1 parameters for synthetic network generation

Parameter	Value
Initial nodes (n_{init})	5
Initial graph model	Erdős-Rényi ($p = 0.5$)
Total snapshots (T)	150
Attachments per new node (m)	Uniform(2, 5)
Detachments per step	Uniform(1, 3)
Discretization bins	5
Bin edges (degree centrality)	[0.00, 0.01, 0.02, 0.03, 0.04, 0.05]
State labels	{Very Low, Low, Medium, High, Very High}
Random seed	42

churn and adaptation. This procedure produces evolving network structures suitable for analyzing centrality trajectories and for probabilistic modeling of influence. Extensions to multilayer networks, where multiple types of edges coexist over the same node set, are also supported to capture multiplex interactions.

For all simulations, parameters such as the initial node count, attachment range, and total time steps are recorded and fixed to ensure reproducibility. This framework provides a flexible and mathematically precise foundation for subsequent analysis of centrality and influence in evolving networks. See Table I for chosen parameters.

B. Dynamic and Adaptive Network Simulation

To study influence and centrality in evolving networks, we extend static network models into dynamic, time-varying structures. Let $\{G_t = (V_t, E_t)\}_{t=1}^T$ denote a sequence of network snapshots over T discrete time steps. Each G_t represents the state of the network at time t , including the current set of nodes V_t and edges E_t .

Simulations begin with a small connected seed graph G_1 , typically generated as an Erdős-Rényi random network to ensure baseline connectivity and reproducibility. At each subsequent time step, network evolution is driven by both growth and adaptation mechanisms. Growth is modeled by the addition of new nodes, where each new node v_{new} connects to a set of m existing nodes sampled according to a specified attachment criterion, which may be uniform, degree-based, or guided by composite measures. This process captures preferential attachment and mimics the emergence of hubs and local leaders over time.

Network adaptation is incorporated through edge addition, deletion, or rewiring. For each node, a small number of edges may be removed or redirected, and new connections may form with nodes in its local neighborhood. Formally, if E_t is the edge set at time t , the updated edge set at $t + 1$ is

$$E_{t+1} = (E_t \setminus E_{\text{del}}) \cup E_{\text{add}},$$

where E_{del} and E_{add} denote edges removed and added, respectively. The rewiring and attachment probabilities can be uniform or weighted by node centrality, trust accumulation, or proximity to influential hubs, allowing the network to adapt in ways that reflect social, biological, or technological dynamics.

At each time step, a snapshot G_t is recorded, enabling the computation of node-level statistics such as degree, betweenness, closeness, and eigenvector centrality. This framework

generates networks that are both temporally coherent and structurally adaptive, forming the foundation for subsequent modeling of centrality trajectories and Markov chain-based influence prediction. By explicitly controlling growth and adaptation parameters, we can systematically investigate how local and global network dynamics affect the evolution of node importance over time.

a) Selection of Centrality Measures: The present study begins its analysis of node importance and prediction using a set of four fundamental and widely studied centrality measures: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. These measures are employed as baseline indicators of node importance, providing interpretable and complementary perspectives on network structure. The utilization of classical centrality measures facilitates a straightforward comparison and serves as a foundational framework for assessing the added value of more complex approaches.

- 1) Degree centrality measures the relative number of connections of a node and captures its immediate local influence. In our implementation, degree centrality is computed using the `nx.degree_centrality()` function from the NetworkX library in Python, which normalizes the node degree by the maximum possible number of connections in the network [15].
- 2) Betweenness centrality quantifies the extent to which a node acts as an intermediary by lying on shortest paths between other node pairs. Nodes with high betweenness centrality often play a crucial role in controlling information flow. We compute this measure using the `nx.betweenness_centrality()` function [17].
- 3) Closeness centrality reflects how efficiently a node can reach all other nodes in the network, based on the average length of shortest paths originating from it. This metric captures the notion of accessibility and speed of information spread and is computed using `nx.closeness_centrality()` [15].
- 4) Eigenvector centrality extends the concept of degree centrality by accounting not only for the number of connections a node has, but also for the importance of its neighbors. Nodes connected to other highly central nodes receive higher scores, capturing a notion of global influence within the network. In our analysis, eigenvector centrality is computed using the `nx.eigenvector_centrality()` function [30].

Collectively, these four centrality measures furnish a balanced and well-established toolkit for analyzing node importance. The computational availability, interpretability, and theoretical significance of these models renders them suitable for initial prediction tasks. In the following sections, we build upon these classical measures by introducing dynamic, probabilistic, and temporal centrality frameworks that are better suited for evolving networks and influence prediction.

C. Using Markov Chains to Predict Centrality

In this section, the focus is on the predictive modeling of node centrality trajectories within dynamic networks. The central idea is to leverage Markov chains as a probabilistic framework to forecast the future state of a node's importance, based on its historical behavior. To this end, an end-to-end experimental pipeline was developed, integrating dynamic network generation, centrality computation, discretizations of continuous measures, transition probability estimation, and predictive evaluation.

a) Discretization of Centrality Trajectories: Since centrality measures are continuous-valued, direct application of Markov chain models is not feasible. To address this, we discretize centrality trajectories into a finite set of qualitative states (e.g. *Very Low, Low, Medium, High, Very High*). This transformation enables categorical analysis and allows the evolution of a node's importance to be interpreted as transitions between symbolic states.

Formally, given a sequence of centrality values $\{c_t\}_{t=1}^T$ for a node over T time steps, we partition the range of values into bins $B = \{b_1, b_2, \dots, b_k\}$. Each observed centrality value is assigned to the corresponding bin, producing a sequence of discrete states $S = \{s_1, s_2, \dots, s_T\}$ where $s_t \in \{1, \dots, k\}$. This sequence serves as the input to our Markov chain model.

To model temporal evolution using Markov chains, we transform continuous centrality values into a small set of discrete states (labels). The default configuration uses 5 bins:

Labels = {Very Low, Low, Medium, High, Very High},

and default bin edges [0.00, 0.01, 0.02, 0.03, 0.04, 0.05]. The discretization function maps a numeric centrality x to the bin index j satisfying $b_j \leq x < b_{j+1}$ (last bin inclusive). Formally, for a centrality sequence $\{x_t\}$ of node i , we define the state sequence

$$s_i = (s_i^1, s_i^2, \dots, s_i^T), \quad s_i^t \in \{\ell_1, \dots, \ell_5\}.$$

Alternative binning strategies are supported:

- Uniform bins on theoretical bounds: compute lower and upper bounds for centralities and set edges with `linspace(min, max, k + 1)`.

b) Markov Chain Modeling and Prediction:

c) First-order Markov model.: From the discretized state sequences $\{s_i\}$, we compute first-order transition counts

$$C_{ab} = \#\{(t, i) : s_i^t = a, s_i^{t+1} = b\},$$

and then transition probabilities

$$T_{ab} = \Pr(s^{t+1} = b \mid s^t = a) = \frac{C_{ab}}{\sum_{b'} C_{ab'}}. \quad (1)$$

To predict the next state given current state a , we either sample from the categorical distribution defined by row T_a . (stochastic prediction), or choose $\arg \max_b T_{ab}$ (deterministic prediction).

d) Markov chain construction.: Using the discretized sequences, we construct a first-order Markov chain for each centrality measure. The transition probability matrix P is estimated empirically as

$$P_{ij} = \frac{N_{ij}}{\sum_j N_{ij}},$$

where N_{ij} denotes the number of observed transitions from state i to state j across all nodes and time steps. Each row of P thus represents a probability distribution over the possible next states, conditioned on the current state. This model encodes the stochastic dynamics of centrality evolution.

e) Centrality Computation and Storage: For each generated or ingested snapshot G_t , we compute the following node-level statistics using NetworkX functions:

$$\text{deg}_t(i), \quad \text{btw}_t(i), \quad \text{clo}_t(i), \quad \text{eig}_t(i).$$

Results are stored as lists of dictionaries (one dictionary per snapshot mapping node to score), e.g. `degree_matrix[t][i] = deg_t(i)`. This sparse mapping reduces memory compared to full matrices when nodes are created over time.

f) Centrality Bounds and Adaptive Binning: The code contains functions to estimate conservative upper and lower bounds for degree and betweenness centralities per node. Bounds can be used to:

- derive dataset-adaptive bin edges (e.g. `linspace(min_bound, max_bound, k + 1)`);
- detect out-of-range centrality values during simulation;
- reason about theoretical limits.

Computing bounds as discussed in Section III-B improves robustness of the discretization step for networks with very different sizes or growth dynamics.

g) Extensibility and Higher-Order Models: While our initial experiments focus on degree centrality and first-order Markov chains, the methodology is readily extensible in two directions:

- 1) additional centrality measures (betweenness, closeness, eigenvector) can be modeled within the same framework;
- 2) higher-order Markov models can be employed to capture memory effects.

h) k-th order Markov model.: To capture richer temporal dependencies, we build a k -th order Markov model by aggregating transition tuples

$$((s^{t-k+1}, \dots, s^t), s^{t+1}).$$

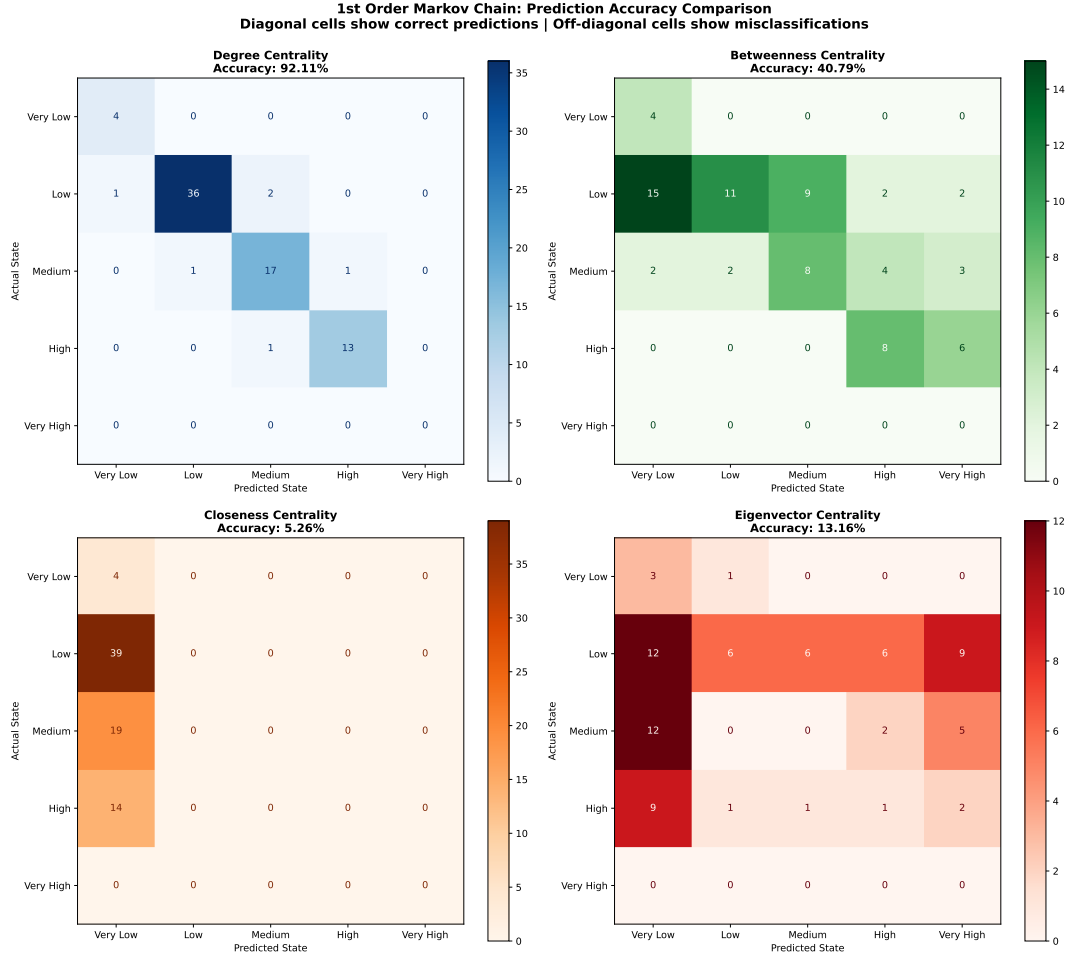


Fig. 1: Confusion matrices for first order Markov predictions of discretized centrality states.

TABLE II: Table showing the evaluation metrics

		F_1	Accuracy
First order Markov Model	DC	0.9589	0.9211
	BC	0.6739	0.5082
	CC	0.1	0.0526
	EC	0.2325	0.1315
Second order Markov Model	DC	0.9792	0.9594
	BC	0.9596	0.9225
	CC	0.9951	0.9904
	EC	0.9699	0.9416

V. EXPERIMENTAL RESULTS

A. First Order Markov Chain Prediction Performance

Prior to the interpretation of the results, it is essential to comprehend the confusion matrix. A confusion matrix is a tabular representation used to evaluate the performance of a classification model by comparing the predicted class labels with the true labels. Each row in the data corresponds to the actual class, while each column represents the predicted class. Predictions that are accurate are indicated by entries along the main diagonal of the matrix, whereas misclassifications are

indicated by off-diagonal entries. The confusion matrix is a data structure that provides a concise overview of the outcomes of a classification process. It facilitates the calculation of performance metrics, including accuracy, precision, recall, and the F_1 -score. As recall was uniformly 1 across all classes and models, Table II reports only accuracy and the F_1 -score.

Each centrality time series was discretized into five bins: Very Low, Low, Medium, High, and Very High. A first-order Markov chain was trained to predict the next state of each node. As illustrated in Figure 1, the resulting confusion matrices are presented.



Fig. 2: Confusion matrices for second order Markov predictions with $k = 2$.

The confusion matrix reveals a strong diagonal dominance, particularly evident in the Low, Medium, and High categories, with respect to degree centrality. This finding suggests that the degree of evolution occurs in a relatively smooth and predictable manner. Misclassifications primarily occur between adjacent bins, such as Low and Very Low or Medium and Low, reflecting gradual changes in connectivity rather than abrupt transitions. As indicated by the findings, betweenness centrality demonstrates a conspicuously diminished diagonal structure, with a considerable proportion of high and medium states being erroneously classified as low. This phenomenon is indicative of the highly volatile nature of betweenness, where nodes can rapidly gain or lose brokerage roles due to the formation or deletion of shortest paths.

Closeness centrality demonstrates a pronounced bias toward the Very Low bin, suggesting that as the network expands, the majority of nodes experience an increase in distance from one another. The model accurately forecasts this collapse, yielding high apparent accuracy but low informational diversity. Eigenvector centrality demonstrates moderate predictability, with substantial ambiguity between Medium, High, and Very High states. This phenomenon underscores the heightened sensitiv-

ity of eigenvector centrality to subtle topological alterations in influential subgraphs.

B. Nodewise Prediction at the Final Timestep

Figure 3 compares the actual and predicted degree centrality bins for all nodes at the final timestep.

The majority of nodes are accurately classified into Low or Medium states, which predominate the distribution at late times. Errors have been found to be concentrated among nodes with borderline values, where minute numerical disparities lead to transitions between adjacent bins. The Markov model has been demonstrated to effectively capture the global stratification of the network, effectively distinguishing between hubs and peripheral nodes, even in large-scale networks.

C. Higher Order Markov Models

The second-order Markov model is predicated on the assumption that two previous states are incorporated to predict the next state. As illustrated in Figure 2 and Table II, the metrics provide a quantitative representation of the data.

Across all centralities, diagonal dominance is found to be significantly stronger than in the first order case. The

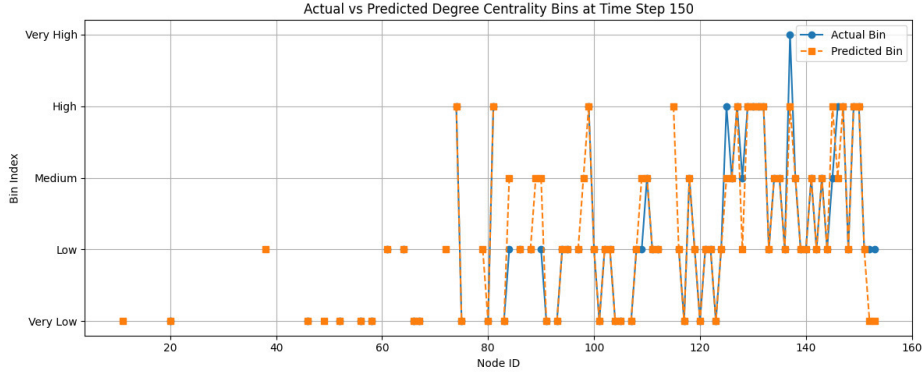


Fig. 3: Actual versus predicted degree centrality bins for all nodes at timestep 150.

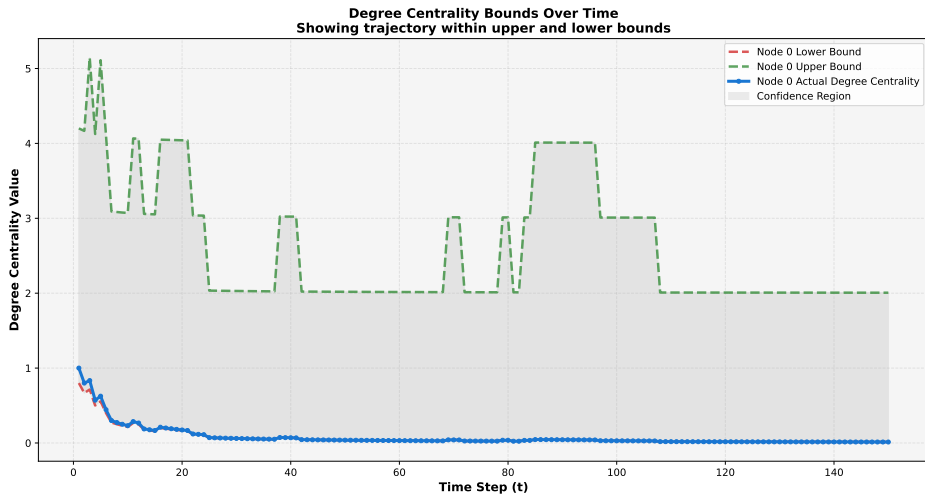


Fig. 4: Degree centrality of Node 0 with theoretical upper and lower bounds over time.

degree and closeness centrality indices achieve near-perfect classification, while the eigenvector and betweenness indices also improve substantially. This finding serves to substantiate the hypothesis that node importance is endowed with short-term temporal memory and that the incorporation of past states significantly enhances predictability.

D. Degree Centrality Bounds and Structural Stability

Figure 4 shows the degree centrality of Node 0, along with analytically derived upper and lower bounds.

The observed degree centrality remains strictly within the theoretical envelope throughout the simulation. In the nascent stages of network evolution, both the upper bound and the actual centrality are high, reflecting hub dominance. As the network expands, both bounds and actual values approach zero, thereby substantiating the notion that the decay in normalized degree is a structural consequence of network growth, rather than a modeling artifact.

E. Betweenness Centrality Bounds

Figure 5 displays betweenness centrality and its theoretical bounds at the final timestep.

All observed values lie far below the maximum theoretical bound, showing that although the network contains hubs, it does not collapse into a single dominating bridge. This reflects the creation of multiple alternative shortest paths as the network grows.

VI. DISCUSSION AND OUTLOOK

The present work has introduced a probabilistic framework for the analysis and prediction of node centrality in evolving networks. In the context of classical network models, the centrality measures—including degree, betweenness, closeness, and eigenvector centrality—were computed at each temporal interval. These measures were subsequently transformed into discrete state trajectories. This discretization enabled the formulation of centrality evolution as a stochastic process, which was modeled using first- and higher-order Markov chains. The transition probability matrices were estimated through empirical means, and the predictive performance was evaluated through the use of classification-based metrics and confusion matrices. The findings indicate that centrality dynamics manifest structured temporal dependencies, and that incorporating

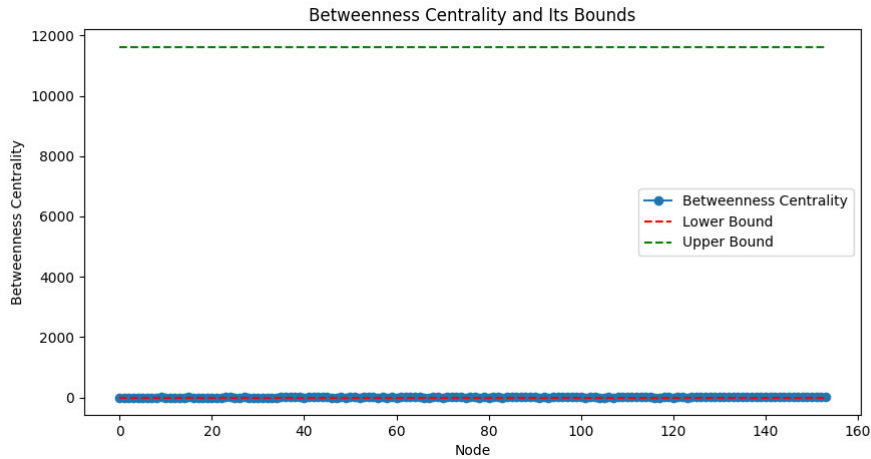


Fig. 5: Betweenness centrality with lower and upper bounds at the final timestep.

short-term memory through higher-order models significantly enhances predictive accuracy across all considered metrics.

A principal strength of the proposed approach is its interpretability. By mapping continuous centrality values to a finite set of states, the framework provides an intuitive description of how node importance evolves, while still retaining a probabilistic foundation. The employment of Markov chains enables the lucid modeling of transition dynamics and expedites the direct interpretation of persistence, volatility, and drift in node influence. The experimental design, predicated on the controlled synthetic network generation, ensures reproducibility and enables systematic investigation of how different topological mechanisms affect centrality evolution. The observed consistency between empirical findings and theoretical predictions, including the limited variability of centrality measures and the gradual progression of degree centrality, further reinforces the approach's validity.

However, it is imperative to acknowledge several limitations inherent to this approach. The discretization of centrality values, while necessary for Markov modeling, introduces a loss of information and may obscure fine-grained variations, particularly near bin boundaries. The model's reliance on first- and second-order Markov assumptions imposes limitations on its capacity to accommodate long-range temporal dependencies, which are likely to be present in complex real-world systems. Furthermore, the simulations are grounded in stylized network models, such as Erdős-Rényi and preferential attachment mechanisms. While these models have been thoroughly validated, they may not fully capture the heterogeneity and domain-specific constraints inherent to empirical networks. Another limitation pertains to the assumption of stationarity in the transition probabilities. However, real networks may exhibit non-stationary dynamics due to external shocks or regime changes. Finally, while the empirical results indicate strong predictive performance, a comprehensive theoretical

analysis of error bounds and convergence properties remains an open problem.

These limitations suggest several promising avenues for future research. One natural extension of this work is the development of non-Markovian or higher-memory models that can capture longer temporal dependencies. These models can be developed through variable-order Markov processes or sequence modeling techniques. Another significant avenue involves the integration of continuous-state models, which would circumvent discretization and enable centrality to be regarded as a fully continuous stochastic process. The incorporation of non-linear relationships between different centrality measures signifies a progression toward realism, as influence in numerous systems emanates from complex interactions rather than linear aggregation. Furthermore, the implementation of the proposed framework in real-world datasets, including social, biological, and technological networks, would offer significant insights into its practical utility and generalizability. Finally, the integration of probabilistic network modeling with contemporary machine learning methodologies, such as graph neural networks, has the potential to unify interpretability with high expressive capacity. This, in turn, promises to enhance our understanding of how influence emerges and evolves in complex dynamic systems.

In summary, the present study demonstrates that centrality is not a static property; rather, it is a dynamic and inherently stochastic quantity that could be effectively modeled and predicted within a probabilistic framework. The work combines network simulation, centrality analysis, and Markov chain modeling to provide a coherent approach to understanding and forecasting node importance. While further theoretical and empirical developments are required, the results highlight the potential of probabilistic methods to advance the study of dynamic networks and to uncover the temporal structure underlying complex systems.

ACKNOWLEDGMENTS

During the preparation of this manuscript, the authors used DeepL to check the language and grammar. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

REFERENCES

- [1] V. Sekara, A. Stopczynski, and S. Lehmann, "Fundamental structures of dynamic social networks," *Proceedings of the national academy of sciences*, vol. 113, no. 36, pp. 9977–9982, 2016.
- [2] J. Dörpinghaus, "Die soziale Netzwerkanalyse: Neue Perspektiven für die Auslegung biblischer Texte?" *Bibisch erneuerte Theologie*, vol. 5, pp. 75–96, 2021.
- [3] F. Grando, L. Z. Granville, and L. C. Lamb, "Machine learning in network centrality measures: Tutorial and outlook," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1–32, 2018.
- [4] J. Dörpinghaus, "Digital theology: New perspectives on interdisciplinary research between the humanities and theology," *Interdisciplinary Journal of Research on Religion*, vol. 18, p. 1, 2022.
- [5] J. Dörpinghaus and M. Jacobs, "Semantic knowledge graph embeddings for biomedical research: Data integration using linked open data." in *SEMANTICS (Posters & Demos)*, 2019.
- [6] J. Dörpinghaus, S. Schaaf, and M. Jacobs, "Soft document clustering using a novel graph covering approach," *BioData mining*, vol. 11, no. 1, p. 11, 2018.
- [7] J. Dörpinghaus, J. Klein, J. Darms, S. Madan, and M. Jacobs, "Scaiview—a semantic search engine for biomedical research utilizing a microservice architecture." in *SEMANTICS (Posters & Demos)*, 2018.
- [8] J. Dörpinghaus, D. Samray, and R. Helmrich, "Challenges of automated identification of access to education and training in germany," *Information*, vol. 14, no. 10, p. 524, 2023.
- [9] M. P. Mangroliya, J. Dörpinghaus, and R. Rockenfeller, "Towards the analysis of errors in centrality measures in perpetuated networks," pp. 417–428. [Online]. Available: https://annals-csis.org/Volume_39/drpf4841.html
- [10] H. T. Tu, T. T. Phan, and K. P. Nguyen, "Modeling information diffusion in social networks with ordinary linear differential equations," *Information Sciences*, vol. 593, pp. 614–636, 2022.
- [11] B. Tadić and R. Melnik, "Self-organised critical dynamics as a key to fundamental features of complexity in physical, biological, and social networks," *Dynamics*, vol. 1, no. 2, pp. 181–197, 2021.
- [12] W.-N. Chen, D.-Z. Tan, Q. Yang, T. Gu, and J. Zhang, "Ant colony optimization for the control of pollutant spreading on social networks," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 4053–4065, 2019.
- [13] F. Menczer, S. Fortunato, and C. A. Davis, *A First Course in Network Science*, 1st ed. Cambridge University Press. [Online]. Available: <https://www.cambridge.org/highereducation/books/first-course-in-network-science/EE22722F27519D8BB1443C7225C57BAF#contents>
- [14] L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press ; BookSurge.
- [15] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [16] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [17] —, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [18] D. F. Gleich and R. A. Rossi, "A Dynamical System for PageRank with Time-Dependent Teleportation," *Internet Mathematics*, vol. 10, no. 1–2, pp. 188–217, Apr. 2014.
- [19] C. E. Graafland, J. M. Gutiérrez, J. M. López, D. Pazó, and M. A. Rodríguez, "The probabilistic backbone of data-driven complex networks: An example in climate," *Scientific Reports*, vol. 10, no. 1, p. 11484, Jul. 2020.
- [20] N. Friedman, K. Murphy, and S. Russell, "Learning the Structure of Dynamic Probabilistic Networks."
- [21] G. Ghoshal and A.-L. Barabási, "Ranking stability and super-stable nodes in complex networks," vol. 2, no. 1, p. 394. [Online]. Available: <https://www.nature.com/articles/ncomms1396>
- [22] M. Ruan, X. Chen, and H. Zhou, "Centrality prediction based on K-order Markov chain in Mobile Social Networks," vol. 12, no. 6, pp. 1662–1672. [Online]. Available: <http://link.springer.com/10.1007/s12083-019-00746-y>
- [23] S. Bartolucci, F. Caccioli, F. Caravelli, and P. Vivo, "Distribution of centrality measures on undirected random networks via the cavity method," *Proceedings of the National Academy of Sciences*, vol. 121, no. 40, p. e2403682121, Oct. 2024.
- [24] *Markov-Chain based centralities and Space Syntax' Angular Analysis: an initial overview and application*, Krakow, Poland, 2023.
- [25] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*. Princeton, NJ: D. Van Nostrand, 1960.
- [26] S. N. Mohammed and S. Gündüç, "TPM: Transition Probability Matrix – Graph Structural Feature based Embedding," Mar. 2023.
- [27] D.-W. Huang and Z.-G. Yu, "Dynamic-Sensitive centrality of nodes in temporal networks," vol. 7, no. 1, p. 41454. [Online]. Available: <https://www.nature.com/articles/srep41454>
- [28] J. Dörpinghaus, V. Weil, C. Düing, and M. W. Sommer, "Centrality measures in multi-layer knowledge graphs," *Annals of Computer Science and Information Systems*, vol. 32, 2022.
- [29] J. Dörpinghaus, V. Weil, and M. W. Sommer, "Towards modeling and analysis of longitudinal social networks: J. dörpinghaus et al." *Applied Network Science*, vol. 9, no. 1, p. 52, 2024.
- [30] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, pp. 55–71, 2005.