Analysis of the Structural Characteristics of Social Networks from a Statistical Perspective

Tianyao Xie
Nanjing Zhonghua High School, Nanjing, Jiangsu, China

Abstract: This paper systematically explores the application research of social network structure from a statistical perspective, focusing on three core areas: modeling of information dissemination dynamics, analysis of social capital accumulation mechanisms, and optimization of network intervention strategies. By integrating multiple statistical methods such as survival analysis, multi-layer network models, and reinforcement learning, the nonlinear influence mechanism of network structure on individual behavior and social systems is revealed. The research emphasizes the innovation of statistical modeling under dynamic network evolution, individual strategy interaction and resource constraints. It combines ABM simulation and Bayesian optimization techniques to construct an experimental verification framework, providing quantitative decision support for real-world scenarios such as social media governance and public health prevention and control, and promoting the transformation of social network analysis from descriptive research to causal inference and optimization design paradigms.

Keywords: Social Network Structure; Statistical Modeling; Dynamics of Information Dissemination; Social Capital; Optimization of Intervention Strategies

1. Introduction

As a carrier of individual interaction and resource flow, the structural characteristics of social networks profoundly influence key social processes such as information diffusion, social capital accumulation and system resilience. Traditional research mostly focuses on the static description of network topology, but neglects the dynamic interaction between structural effects and individual behaviors. With the development of big data and computational statistics methods, analyzing the generation mechanism, evolution laws and intervention paths of network structures from a statistical perspective has become an important frontier. This paper systematically reviews the application research of social network structure in three major fields: information dissemination, social capital and intervention strategies. By integrating complex network theory, statistical inference and optimization algorithms, it reveals the deep logic of network-driven social phenomena and provides theoretical support for the construction of a data-driven social governance paradigm.

2. Theoretical Framework for Statistical Modeling of Social Network Structure

2.1 Construction of the Statistical Description Index System for Network Topology

In the theoretical framework of statistical modeling of social network structure, the construction of the statistical description index system of network topology is a fundamental task. Its complexity is not only reflected in the diversity of index dimensions, but also in the potential nonlinear correlations and hierarchical nested relationships among different indicators[1-2]. Traditional research has mostly focused on the independent analysis of single centrality measures (such as degree centrality and betweenness centrality), but the multi-layer coupling characteristics of modern complex networks require the construction of a more explanatory composite index system: On the one hand, it is necessary to integrate local structural features (such as the clustering coefficient of node neighborhoods) and global topological attributes (such as the average path length of the network), and achieve dimensionality reduction and fusion through principal component analysis or structural equation models. On the other hand, a dynamic weight mechanism needs to be introduced to weighted and coupled the heterogeneity of node attributes (such as individual social capital and information dissemination capacity) with the importance of structural position, forming a comprehensive measure that reflects the synergy of "structure-attribute". It is worth noting that this indicator system also needs to meet the requirements of statistical testability. The independence between indicators should be verified through QAP correlation analysis, and the stability of the measure should be evaluated by using bootstrap resampling technology. Furthermore, in the face of multi-layer network scenarios, the mapping transformation and consistency calibration of cross-layer indicators have become new challenges. There is an urgent need to develop multi-dimensional indicator construction methods based on tensor decomposition to achieve a leap in the statistical description paradigm from planar networks to three-dimensional topologies. This process not only requires rigorous mathematical derivation support but also relies on the verification and optimization of large-scale network simulation experiments, ultimately forming a statistical description framework that is both theoretically complete and practically operable.

2.2 Statistical Inference of Complex Network Generation Mechanisms

In the theoretical framework of statistical modeling of social network structure, the statistical inference of the generation mechanism of complex networks constitutes the core link connecting the micro-interaction rules and the macro-topological emergence. Its complexity stems from the interwoven effects of multiple heterogeneous factors during the network evolution process[3-4]. Most existing studies are based on ERGM (Exponential Random Graph Model) or PA (Preference Connection) mechanisms, but the generation of real networks often has the dual characteristics of random disturbances and deterministic trends: On the one hand, a hierarchical generative model needs to be constructed through the Bayesian non-parametric method to decoupled the inherent attributes of nodes (such as social status and resource endowment) and dynamic behavior patterns (such as strategy selection and information transmission) into latent variables, and then the MCMC algorithm is utilized to achieve adaptive exploration of the parameter space. On the other hand, a time-varying parameter mechanism needs to be introduced to capture the non-stationarity of network evolution, such as depicting the phased mutations of connection probabilities through state space models, or using Gaussian process regression to fit the time-dependent structure of clustering coefficients. It is particularly crucial that the validity of statistical inference highly depends on the rigor of hypothesis testing-not only is it necessary to design a network generation process identifiability verification based on permutation testing to prevent false associations caused by model missetting, but also to develop a multi-model comparison framework (such as the combined application of BIC criteria and WAIC criteria) to identify the relative contribution of different generation mechanisms (such as triangular closure motivation and homogeneity preference). This statistical inversion from micro rules to macro models is essentially a quantitative interpretation of the self-organizing characteristics of complex systems. Its methodological breakthrough will provide a new paradigm for understanding the structural resilience of social networks.

2.3 Statistical Time Series Analysis of Dynamic Network Evolution

In the theoretical framework of statistical modeling of social network structure, the statistical time series analysis of dynamic network evolution, as the core method for capturing the time-varying characteristics of network structure, its complexity stems from the profound contradiction between the nonlinear dynamics of network state transitions and the sparsity and noise pollution of observed data[5-6]. Traditional studies mostly adopt the sliding window method to extract temporal features, but the fixed window scale is difficult to adapt to the multifractal characteristics of network evolution. There is an urgent need to construct an adaptive variable window model-for example, dynamically optimizing the window width through information criteria, or using wavelet transform to achieve time-frequency joint analysis to capture the coexistence of abrupt points and periodic patterns. Furthermore, the probabilistic modeling of network state transitions needs to break through the limitations of the Markov hypothesis and introduce high-order memory effects: a deep learning-based LSTM-GRU hybrid neural network can be constructed to encode the historical dependency of node connections as hidden state vectors, and at the same time, the contribution weights of different time steps to the current structure can be quantified through the attention mechanism. Semi-parametric methods can also be developed, such as modeling the evolution of network density as a threshold autoregressive model with structural mutations, and using Bayesian change point detection algorithms to identify key turning points. It is worth noting that the statistical validity of dynamic network analysis highly depends on the missing data handling strategy-an iterative algorithm based on matrix completion needs to be designed to fill the time series faults while preserving network sparsity, or to generate pseudo-observation sequences through multiple imputation to correct estimation bias. This process essentially transforms the self-organizing evolution laws of complex systems into computable statistical language. Its methodological innovation will provide a dynamic perspective for understanding the formation and collapse mechanisms of social network resilience.

3. Empirical Statistical Methods for the Structural Characteristics of Social Networks

3.1 Statistical Preprocessing Techniques for Multi-Source Heterogeneous Network Data

In the empirical statistical method system of social network structure characteristics, the statistical preprocessing technology of multi-source heterogeneous network data serves as a key bridge connecting raw data and structural analysis. Its complexity stems from the multi-dimensional heterogeneity of network data at the source, format and semantic levels[6-7]. For network data collected from different channels such as social media, organizational relationships, and communication records, a hierarchical preprocessing framework needs to be constructed: Firstly, at the data fusion level, it is necessary to develop ontology-based semantic alignment algorithms. By defining a unified meta-model of node attributes and edge relationships, heterogeneous networks can be mapped to the isomorphic feature space. Meanwhile, the embedding technology of graph neural networks is utilized to achieve dimensionality reduction and fusion of cross-source structural features. Secondly, in the noise cleaning stage, a dual filtering mechanism combining statistical tests and domain knowledge needs to be designed-it not only uses chi-square tests to identify low-frequency but significant structural patterns (such as rare but critical bridge nodes), but also employs cluster analysis to eliminate abnormal subgraphs caused by data acquisition errors. Finally, in view of the data sparsity problem specific to dynamic networks, it is necessary to innovate the interpolation method for missing values: A spatio-temporal joint model based on matrix factorization can be constructed to encode the temporal dependence of network snapshots and the spatial correlation of node attributes as low-rank constraints, or to synthesize pseudo-observation data through a generative adversarial network (GAN), thereby enhancing data density while retaining the original network topological features. This process is essentially a multi-dimensional "purification" of complex network data, and its methodological breakthrough will lay a reliable data foundation for the subsequent extraction of structural features.

3.2 Statistical Significance Test System for Network Structure Characteristics

In the empirical statistical method system of social network structure characteristics, the statistical significance test system of network structure characteristics, as the core tool for verifying theoretical hypotheses, its complexity stems from the dual challenges of non-independence and high dimensionality unique to network data. Traditional statistical test methods (such as t-test and chi-square test) often lead to the expansion of the first type of error when applied to network data due to the neglect of the dependency relationship between nodes. There is an urgent need to construct an improved framework based on permutation test: By designing an edge exchange algorithm that retains the topological attributes of the network, a random network distribution that conforms to the null hypothesis is generated, and then the empirical p-value of the empirical eigenvalue is calculated. This is particularly suitable for the verification of local structural features such as community division and centrality measure. For global features (such as network density and average path length), it is necessary to develop a Bayesian test method based on Markov Chain Monte Carlo (MCMC), combining the posterior distribution of structural features with prior knowledge, and quantifying the strength of evidence by calculating Bayesian factors, effectively overcome the problem of low test efficiency caused by insufficient sample size.

What is more complex is that in multi-layer network scenarios, cross-layer joint test statistics need to be constructed. For instance, the common patterns of each layer can be extracted through tensor decomposition, and then the Copula function can be used to model the inter-layer dependency structure, achieving a paradigm upgrade from planar test to three-dimensional inference. The improvement of this verification system not only requires innovative breakthroughs in statistical theory, but also needs to be combined with large-scale network simulation experiments to verify its robustness, ultimately forming a network structure feature verification framework that combines theoretical rigor with practical operability.

3.3 Design of Statistical Simulation Experiments for Network Robustness

In the empirical statistical method system of social network structure characteristics, the statistical simulation experiment design of network robustness serves as a key path for evaluating the system's anti-interference ability. Its complexity stems from the dual coupling of network structure dynamics and the heterogeneity of disturbance types[8-9]. To address the differentiated impact mechanisms of random attacks and targeted attacks, a hierarchical simulation framework needs to be constructed: In the low-level perturbation generation stage, it is necessary not only to model the random failure of nodes/edges through the Poisson process, but also to develop targeted attack algorithms based on node centrality sorting (such as deleting key nodes by degree centrality and betweenness centrality gradients), and at the same time introduce adaptive perturbation strategies for example, using reinforcement learning to dynamically adjust the attack path to simulate the strategy optimization behavior of attackers in real scenarios. In terms of the response modeling of middle-level networks, it is necessary to integrate percolation theory and survival analysis methods: on the one hand, by monitoring the variation trajectory of the scale of the maximum connected subgraph, piecewise regression is used to identify the critical threshold of network collapse; On the other hand, the node survival time is modeled as a Cox proportional hazards model to quantify the nonlinear moderating effect of structural attributes (such as clustering coefficients and modularity) on robustness. It is particularly crucial that the top-level experimental design achieve a dual rigor of "control variables-scenario expansion"-it is necessary to control confounding factors such as network scale and density through Latin square design, and also to expand to complex scenarios such as multi-layer networks and time-varying networks, for example, to simulate cascading failures caused by inter-layer dependency breaks in multi-layer coupled networks. The construction of this experimental system is essentially to transform the theoretical assumptions of network science into falsifiable statistical inferences. Its methodological innovation will provide quantitative experimental support for understanding the structural resilience of social networks.

4. Application Research on Social Network Structure from a Statistical Perspective

4.1 Statistical Modeling and Prediction of Information Dissemination Dynamics

In the application research of social network structure from a statistical perspective, the statistical modeling and prediction of information dissemination dynamics, as the core direction for revealing the laws of information diffusion in social networks, its complexity stems from the deep interweaving of individual behavior heterogeneity and the dynamics of network structure during the information dissemination process. Although the traditional SIR (Susceptibility-Infection-Recovery) model can depict the basic outline of information dissemination, it is difficult to capture the nonlinear decision-making mechanism of users' forwarding behavior. It is necessary to construct a mixed-effects model based on survival analysis, incorporating user attributes (such as authority, interest preferences) and network locations (such as centrality, structural holes) into random effect terms. At the same time, the Cox proportional hazards model is used to quantify the non-proportional impact of information content characteristics (such as emotional tendency and topic popularity) on the dissemination rate. In view of the time-varying characteristics of the propagation network, it is necessary to develop a dynamic Bayesian network model: the time-varying parameters are estimated through the Markov Chain Monte Carlo (MCMC) algorithm, and the transition points of the propagation stage are identified in combination with the Hidden Markov Model (HMM), and then the dynamic prediction of the propagation path is achieved by using particle filter technology. What is particularly crucial is to build a statistical verification framework for multi-source data fusion-it is necessary to compare the prediction accuracy of different models in real communication scenarios through A/B testing, and also to use social robot detection algorithms to eliminate the interference of automated accounts on the communication pattern. Ultimately, a statistical modeling system with both theoretical explanatory power and practical application value should be formed provide quantitative decision support for public opinion monitoring and information regulation on social media platforms.

4.2 Analysis of the Network Statistical Mechanism of Social Capital Accumulation

In the application research of social network structure from a statistical perspective, the analysis of the network statistical mechanism of social capital accumulation, as a key proposition to reveal the acquisition path of individual social resources, its complexity stems from the dynamic nonlinear system formed by the interaction between network structure effects and individual behavioral strategies. Most existing studies focus on the direct impact of structural positions (such as centrality and structural holes) on social capital, but ignore the moderating role of individual strategic interactions in the process of network evolution-it is necessary to construct a spatial econometric model based on multi-layer networks. The direct connection relationships among individuals (such as strong relationships) and indirect structural embedments (such as weak relationship Bridges) are deconstructed into spatial weight matrices at different levels. Meanwhile, a dynamic game theory framework is introduced to quantify the decision-making process by which individuals achieve the appreciation of social capital through relationship investment (such as time input and emotional support) and strategy adjustment (such as choosing homogeneous or heterogeneous connections). For the heterogeneous dimensions of social capital (such as cognitive capital and relational capital), it is necessary to develop a hybrid method of quantile regression and latent category analysis: through quantile regression to identify the asymmetry of network structure effects at different capital levels, and combined with latent category analysis to reveal the role differentiation of individuals in the network (such as the difference in capital accumulation paths between core nodes and edge nodes). What is particularly crucial is to construct an intertemporal statistical verification framework-using the instrumental variable method to address the endogeneity problem caused by network self-selection, and combining the panel data of social networks to analyze the long-term dynamic effects of capital accumulation, ultimately forming a three-dimensional statistical interpretation of the mechanism of social capital accumulation driven by the network.

4.3 Statistical Optimization Design of Network Intervention Strategies

In the application research of social network structure from a statistical perspective, the statistical optimization design of network intervention strategies, as the core means to enhance the governance efficiency of social systems, its complexity stems from the multi-dimensional coupling of network dynamic evolution, individual heterogeneous responses, and resource constraints. Traditional intervention strategies mostly select target nodes based on static network centrality indicators (such as degree centrality and mediation centrality), but ignore the cascade effect of intervention behaviors and individual strategic adaptation. It is necessary to construct a dynamic optimization framework based on reinforcement learning and model the evolution of network states as a Markov decision process. The intervention strategies (such as information placement and resource allocation) are dynamically adjusted through the Deep Q-Network (DQN) algorithm, and at the same time, the game theory module is introduced to quantify the behavioral feedback of individuals after perceptual intervention (such as cooperation, betrayal or imitation). For multi-objective intervention scenarios (such as simultaneously optimizing information dissemination efficiency and social cohesion), it is necessary to develop multi-objective optimization algorithms: using Pareto frontier analysis to identify the non-dominated solution set of strategy combinations, combining the Analytic Hierarchy Process (AHP) to determine the objective weights, and then searching for the optimal intervention path through genetic algorithms. What is particularly crucial is to establish a statistical experimental verification system-using ABM (Agent-based Modeling) technology to simulate the network evolution trajectories under different strategies, combining Bayesian optimization algorithms for efficient exploration of the parameter space, and ultimately quantifying the causal effects of intervention strategies through a counterfactual reasoning framework Provide statistical optimization solutions that are both theoretically rigorous and practically operable for real-world scenarios such as public health prevention and control and social media governance.

5. Conclusions

This article systematically expounds the paradigm transformation of social network structure research from two dimensions: statistical method innovation and application scenario expansion. By constructing dynamic Bayesian networks, spatial econometric models and reinforcement learning optimization frameworks, the limitations of traditional static analysis have been broken through, and the joint modeling of network evolution, individual decision-making and system response has been achieved. Research has confirmed that the accumulation of social capital exhibits dual path dependence characteristics of structural holes and strong relationships, while the effectiveness of network intervention is significantly constrained by the adaptability of individual strategies and the fairness of resource allocation. Future research needs to further integrate multimodal data and causal inference techniques to explore new forms and governance paths of network structures in the digital society, providing a scientific basis for addressing the challenges of social complexity in the context of globalization and intelligence.

References

[1] Xue Q. Modeling and A/B Testing Based on Social Network Data [D]. Kunming University of Science and Technology,2023.
[2] Lu Q, Ning J. Analysis of the Network Structure of the Beijing-Tianjin-Hebei Urban Agglomeration Based on Enterprise Data: Application of Regional Social Network and Hypernetwork Analysis[J]. Urban Planning 2024,48(9):19-30.
[3] Wang Z, Xu Y. Analysis of the Structural Characteristics and Effects of Spatial Correlation Network in China's Digital Economy: Based on Social Network Analysis [J]. Finance and Economics 2022(10):14.
[4] Zhang X, Han H, Tang Y, et al. Research on the Structural Characteristics of Chinese Urban Network Based on Baidu Migration Data [J]. Journal of Geoinformation Science 2021,23(10):11.
[5] Wang Y, Huang X. The Network Structure and Spatial Correlation of Chinese Cities: An Analysis from "Network Flow" Data [J]. Finance Theory and Practice 2021, 1,042(002):112-118.
[6] Bao Haiqin, Zhang Andong, Sun Weiwei. Social Network Characteristics of International Mobility of Engineering Postgraduates: An Analysis Based on the Resumes of Authors from Three Journals Indexed by IEEE [J] Higher Education Research, 2024, 45(4):98-109.
[7] Jing Meng, Liu Guizhen, Li Qi, et al. Research on the Risk Transmission Characteristics of Carbon Dioxide Geological Storage Based on Social Network Analysis Method [J] Journal of Geology of Chinese Universities, 2023, 29(1):100-109.
[8] Zhang Xuhong, Zhou Cheng, Jin Yiting, et al. Evolution Characteristics and Influencing Mechanism of Urban Digital Economy Network in the Yellow River Basin [J]. Economic Geography, 2025, 45(5):35-45.
[9] Gan Xiaolong, Xiang Shuang, Yi Yan, et al. Analysis of the Policy Evolution Characteristics and Development Trends of China's Digital Rural Construction from the Perspective of Social Network Analysis [J] Journal of Southwest University (Natural Science Edition), 2025, 47(1):147-162.