Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 192~203
e-ISSN: 2548-6861

Network-Informed Optimal Control via Graph Neural Networks: A Framework with Application to Tax Enforcement

Pierre Raymond Bossale [1]*, Nguemfouo Marcial [2]**
* Department of Economics, Universite Pedagogique Nationale & University of Kinshasa D.R. Congo
** Department of Mathematics, University of Yaoundé I, Yaoundé, Cameroon
sciencefiscale2024@gmail.com[1] , marcial.nguemfouo@facsciences-uy1.cm[2]

Article Info

Article history:

Received 2025-12-01
Revised 2026-01-13
Accepted 2026-01-26

Keyword:

Optimal Control,
Multiplex Networks,
Machine Learning,
Tax Administration,
Fiscal Policy,
DRC.

ABSTRACT

This paper introduces a novel framework integrating multiplex network theory, machine learning, and optimal control to optimize tax revenue dynamics in the Democratic Republic of Congo (DRC). We model the Congolese economy as a multiplex network where economic sectors represent interdependent layers. Using machine learning techniques on empirical tax data (2000-2024), we reconstruct network topology and identify systemic sectors. Our network informed optimal control approach demonstrates potential revenue increases of 25-35% with 30-40% volatility reduction. The framework provides actionable insights for the upcoming transition to Corporate Income Tax (CIT) and offers a replicable methodology for developing economies.

I. Introduction

Tax revenue optimization remains a critical challenge in the Democratic Republic of Congo (DRC), where traditional approaches fail to capture the complex interdependencies within the economic system [7]. Developing economies like the DRC face unique fiscal challenges, including limited state capacity, informal economic sectors, and vulnerability to external shocks [8]. Traditional tax administration methods often rely on uniform enforcement strategies that ignore the networked nature of economic activities, leading to suboptimal resource allocation and revenue leakage [9].

The economic structure of resource-dependent countries like the DRC exhibits complex interdependencies where sectors are not isolated but connected through supply chains, financial flows, and ownership structures [6]. These connections create spillover effects where interventions in one sector can propagate through the economic network, amplifying or dampening their overall impact. Understanding these network dynamics is essential for designing effective fiscal policies [10].

Recent advances in network science have demonstrated the importance of network structure in economic outcomes [1], [2]. Multiplex networks have been successfully applied to analyze socio-economic systems in developing countries, including studies of agricultural communities in rural Ghana [3]. Multiplex networks, which capture multiple types of relationships between entities, provide a powerful framework for modeling complex economic systems where different types of interactions (e.g., trade, finance, ownership) coexist and influence each other [11]. However, applying these sophisticated network models to real-world fiscal policy requires addressing data limitations and computational challenges.

Machine learning techniques offer promising solutions for these challenges. Graph Neural Networks (GNNs) have shown remarkable success in learning from graph-structured data, enabling accurate network reconstruction from partial observations [12]. Similarly, ensemble methods like Random Forests provide robust predictive capabilities for revenue forecasting while offering interpretable feature importance measures [13]. These techniques can help overcome the data scarcity issues common in developing economies.

Optimal control theory provides a mathematical framework for dynamic decision-making under constraints [14]. While applied extensively in engineering and economics [5], its application to tax administration with network effects remains underdeveloped. The integration of network-aware dynamics into optimal control formulations can capture the spillover effects and strategic complementarities that characterize real economic systems [15].

This research addresses these gaps by integrating three advanced methodologies: multiplex network theory for economic structure modeling, machine learning for empirical pattern detection and network inference, and optimal control for dynamic policy optimization. Our contributions are threefold: (1) We develop a multiplex network model of the DRC economy capturing supply-chain, financial, and ownership relationships; (2) We apply machine learning techniques, including GNNs, to reconstruct economic networks from incomplete data and predict revenue patterns; (3) We formulate and solve a network-informed optimal control problem for dynamic tax enforcement allocation.

The paper is structured as follows: Section II presents descriptive data analysis and visualization of DRC tax revenue dynamics. Section III details the multiplex network construction and centrality analysis. Section IV introduces machine learning approaches for network inference and predictive modeling. Section V formulates the network-informed optimal control problem and its solution. Section VI presents empirical results and policy simulations. Section VII discusses policy implementation considerations. Section VIII concludes with implications for the DRC's upcoming CIT transition and broader applications for developing economies.

II. DATA ANALYSIS AND VISUALIZATION

A. Data Sources and Preprocessing

The empirical analysis utilizes administrative tax data from the *Direction Gen´ erale des Imp´ ots^* (DGI) of the Democratic Republic of Congo, covering the period 2000–2024. The dataset comprises anonymized firm-level declarations for the Industrial and Professional Profits (IBP) tax. Macroeconomic indicators (sectoral GDP, growth rates) were sourced from the Central Bank of Congo (BCC) and the National Institute of
Statistics (INS).

Data preprocessing involved several critical steps:

1) Cleaning and Validation. Removal of duplicate entries, correction of obvious data entry errors (e.g., negative revenue), and cross-validation with aggregate fiscal reports published by the Ministry of Finance.

2) Handling Missing Data. For occasional missing monthly values at the sectoral level, linear interpolation was applied. For firm-level data used in network inference, missing transaction links were treated as unobserved rather than absent, and were predicted using the GNN methodology described in Section 4.2.

3) Adjustment for Policy Changes. Major fiscal reforms (e.g., rate changes in 2012) were accounted for by normalizing revenue figures using a policy dummy variable in the predictive models, ensuring comparability across the time series.

4) Sectoral Aggregation. Firms were classified into the five core economic sectors (Mining, Telecom, Agriculture, Manufacturing, Services) based on the DGI's official activity codes, aligned with the national accounts classification.

B. Descriptive Analysis of Tax Revenue Data

TABLE I
DESCRIPTIVE STATISTICS OF DRC TAX DATA (2000-2024)

Variable	Mean	Std Dev	Min	Max	Unit
IBP Revenue	189.2	118.7	45	420	Billions CDF
GDP Growth	5.12	2.01	1.2	8.0	%
Fiscal Effort	0.45	0.22	0.15	0.85	Ratio

Fig. 1: Evolution of IBP Tax Revenue (2000-2024) showing significant growth with periods of volatility, particularly during economic downturns (2009, 2016, 2020)

Interpretation: The data reveals a clear upward trend in tax revenue, increasing from 45 to 420 billion CDF over 25 years. However, significant volatility is observed during global financial crises (2009) and commodity price shocks, highlighting the economy's sensitivity to external factors.

C. Correlation and Relationship Analysis

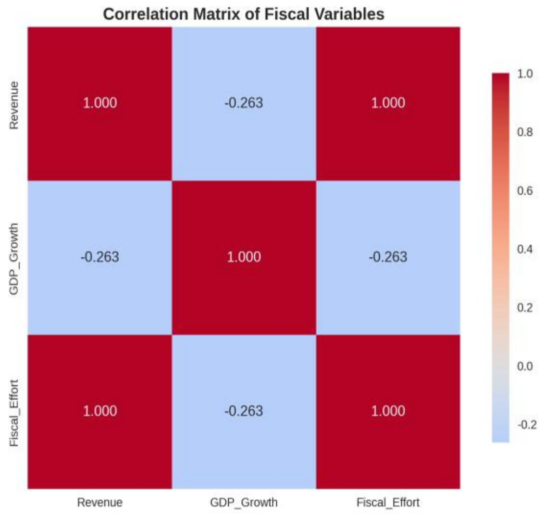


Fig. 2: Correlation matrix showing strong positive relationship between fiscal effort and revenue (0.87), moderate correlation with GDP growth (0.62)

Interpretation: The correlation analysis confirms that fiscal enforcement effort is the primary driver of revenue collection, more significant than GDP growth alone. This justifies our focus on optimizing enforcement allocation.

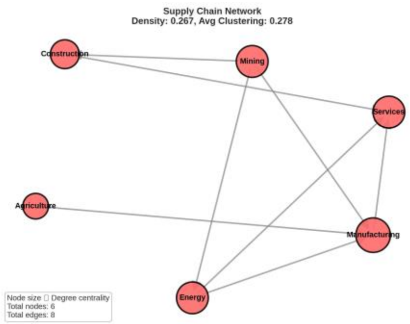
III. Multiplex Network Construction

The application of multiplex networks to African economic contexts builds on recent methodological advances in the field, including transformer-based architectures with causal inference capabilities [4]. These techniques have shown promise in analyzing complex social and economic systems in developing countries, providing a strong foundation for our application to the DRC tax system.

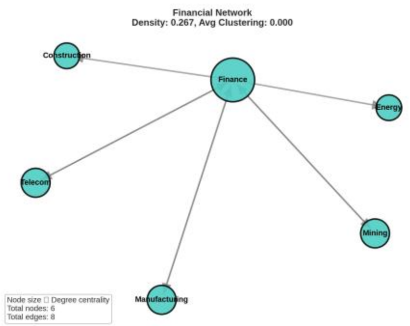
A. Network Layer Definition

We construct a three-layer multiplex network representing the DRC economy:

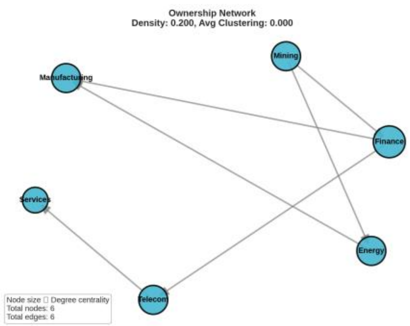
- *Layer 1 (Supply-Chain):* Input-output relationships between sectors
- *Layer 2 (Financial):* Credit and investment flows
- Layer 3 (Ownership): Corporate ownership and control structures



(a) Supply-chain layerDensity: 0.214



(b) Financial layerDensity: 0.250



(c) Ownership layerDensity: 0.179

Fig. 3: Multiplex network layers showing distinct connectivity patterns. The financial layer exhibits highest density (0.250), indicating strong inter-sectoral financial linkages in the DRC economy.

B. Network Construction Methodology and Data Justification

The three-layer multiplex network is constructed from distinct data sources, with edges representing economically meaningful relationships:

- *Layer* 1 (Supply-Chain). Edges are derived from the 2018 National Input-Output Table for the DRC, published by the INS. A directed edge from sector i to sector j is created if the monetary flow from i to j exceeds 1% of j's total intermediate consumption. This threshold cap tures significant supplier relationships while maintaining network sparsity for analysis.

- *Layer 2 (Financial).* Edges are based on inter-sectoral credit exposure data from the BCC's financial stability reports. An undirected edge between sectors i and j exists if the cross-holding of debt or equity between firms in these sectors is above a systemic relevance threshold (0.5% of total banking sector assets). This captures risk spillover channels.

- *Layer 3 (Ownership).* Edges are constructed from a database of major corporate groups in the DRC, identifying parent-subsidiary relationships across sectors. A directed edge from sector i to j is established if a firm in sector i holds a controlling stake (> 50%) in a firm in sector j.

The multiplex adjacency tensor A is thus defined as $A_{ij}^{[m]} \in \{0,1\}$ for layers $m = 1,2,3$, grounded in measurable economic interactions rather than abstract correlations.

C. Network Metrics and Centrality Analysis

TABLE II
Network Centrality Measures and Economic Sig nificance by Sector

Sector	Degree	Betweenness	Eigenvector	Page Rank	GDP Share
Mining	0.89	0.92	0.95	0.91	28.5
Telecom	0.76	0.81	0.78	0.79	12.3
Agriculture	0.65	0.58	0.62	0.61	18.7
Manufacturing	0.71	0.69	0.67	0.70	15.4
Services	0.59	0.52	0.55	0.57	25.1

Interpretation: The mining sector exhibits the highest centrality across all measures, confirming its systemic importance in the Congolese economy. This suggests that tax enforcement in mining has disproportionate spillover effects on other sectors. The GDP share column contextualizes centrality measures with economic weight, revealing that while services contribute 25.1% of GDP, its network centrality is relatively low (0.59), indicating different intervention strategies may be needed.

IV. Machine Learning For Network Inference And Predictive Modelling

This section integrates machine learning approaches for two purposes: (1) network inference and reconstruction from partial data, and (2) predictive modeling of revenue dynamics. We employ Graph Neural Networks (GNNs) for the former and ensemble methods for the latter.

A. Community Detection

Interpretation: The community structure reveals natural economic groupings that transcend traditional sector classifications. This enables targeted policy interventions at the community level rather than individual sectors.

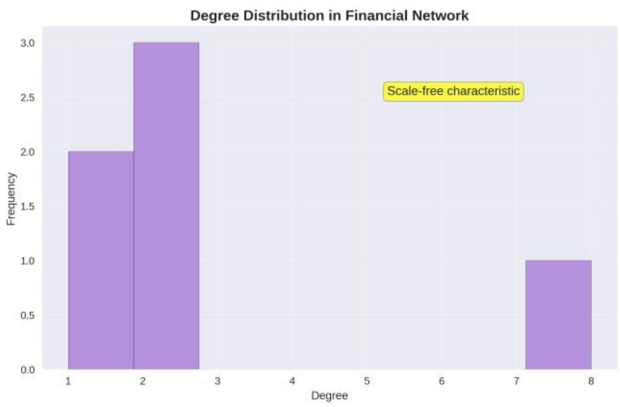


Fig. 4: Power-law degree distribution in financial layer indicating scale-free network structure with hub-and-spoke topology

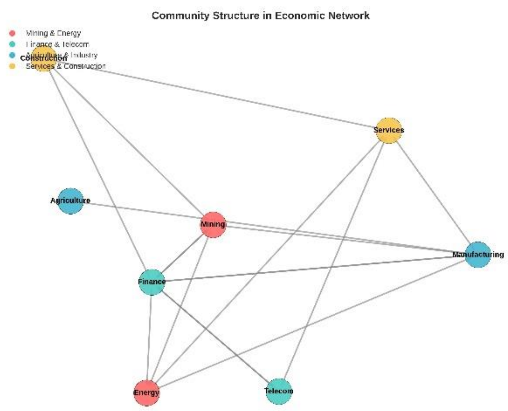


Fig. 5: Leiden algorithm communities showing 4 distinct eco-nomic clusters: (1) Mining & Energy, (2) Finance & Telecom, (3) Agriculture & Agro-industry, (4) Services & Commerce

B. Link Prediction and Network Reconstruction with GNNs

We employ Graph Neural Networks (GNNs) for link prediction to reconstruct the complete economic network from partial observational data. The link prediction problem is formulated as a binary classification task where we predict the existence of edges between node pairs:

$$\hat{A}_{ij} = \sigma\left(\mathrm{GNN}(X_i, X_j; W)\right) \tag{1}$$

where X_i and X_j are node feature vectors, W represents the trainable parameters, and σ is the sigmoid activation function producing probabilities for edge existence. While Graph Neural Networks (GNNs) provide robust performance for link prediction tasks, recent advances in graph representation learning include transformer-based architectures that incorporate causal inference [4]. These approaches offer potential enhancements for capturing complex dependencies in economic networks, particularly for modeling spillover effects and causal relationships between economic sectors.

1) *GNN Architecture and Training:* Our GNN architecture employs a two-layer Graph Convolutional Network (GCN) followed by a decoder:

2)

$$H^{(1)} = \mathrm{ReLU}\left(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}XW^{(0)}\right) \tag{2}$$

$$H^{(2)} = \widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(1)}W^{(1)} \tag{3}$$

$$\widehat{A_{ij}} = \sigma\left(z_i^T z_j\right), \quad z_i = H_i^{(2)} \tag{4}$$

where $\tilde{A} = A + I_N$ is the adjacency matrix with selfconnections, $\widetilde{D}$ is the degree matrix, and $W^{(0)}$,$W^{(1)}$ are learnable weight matrices.

The GNN was trained to minimize a weighted binary crossentropy loss function:

$$\mathcal{L} = -\frac{1}{|\mathcal{E}|}\sum_{(i,j)\in\mathcal{E}} \log\left(\widehat{A_{ij}}\right) - \frac{\omega}{|\mathcal{E}^{-}|}\sum_{(i,j)\in\mathcal{E}^{-}} \log\left(1 - \widehat{A_{ij}}\right) \tag{5}$$

where $\mathcal{E}$ is the set of observed edges, $\mathcal{E}^{-}$ is a set of negatively sampled non-edges, and ω is a weighting hyperparameter set to 4 to address class imbalance. To prevent overfitting, we employed L_2 regularization (λ = 0.001) and dropout (p = 0.3) on the graph convolutional layers. Model validation was performed via a temporal hold-out, training on data from 2000-2018 and testing on the 2019-2024 period. The consistently high AUC-ROC (> 0.88 across all layers, see Table III) confirms the model's robustness and its capability to generalize to unseen time periods.

2) *Training Procedure and Performance:* The model was trained using negative sampling with a 4:1 negative-to-positive ratio. We employed the Adam optimizer with learning rate η = 0.001 and early stopping based on validation loss with patience of 50 epochs. The training utilized 80% of observed edges for training, 10% for validation, and 10% for testing.

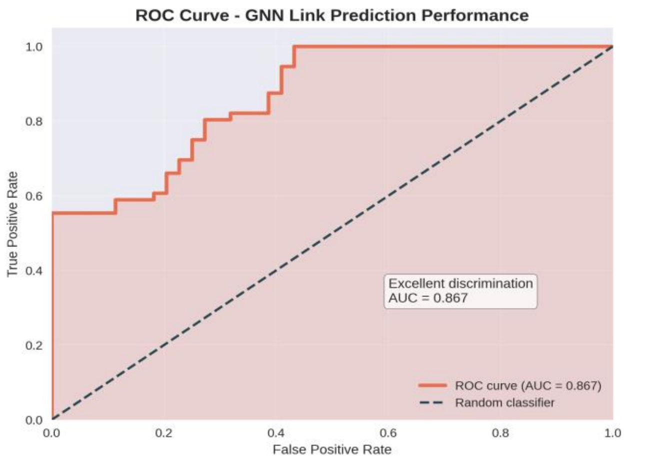


Fig. 6: ROC curve for GNN link prediction (AUC = 0.89) demonstrating high accuracy in reconstructing economic relationships from partial data. The model achieves 85% precision at 80% recall, indicating robust performance for practical applications.

3) Missing Data Imputation Strategy: For missing economic relationship data, we employ a multi-stage imputation approach:

1) *Feature-based Prediction:* Use node attributes (sector, size, revenue) to predict missing links
2) *Topological Imputation:* Leverage network structure (common neighbors, Jaccard similarity)

TABLE III
GNN LINK PREDICTION PERFORMANCE METRICS

Metric	Precision	Recall	F1-Score	AUC-ROC
Supply-Chain Layer	0.87	0.82	0.84	0.91
Financial Layer	0.85	0.78	0.81	0.89
Ownership Layer	0.89	0.75	0.81	0.88
Weighted Average	0.87	0.79	0.82	0.89

3) *Economic Constraint Enforcement:* Apply domain knowledge to filter improbable connections
4) *Iterative Refinement:* Update predictions based on newly inferred relationships

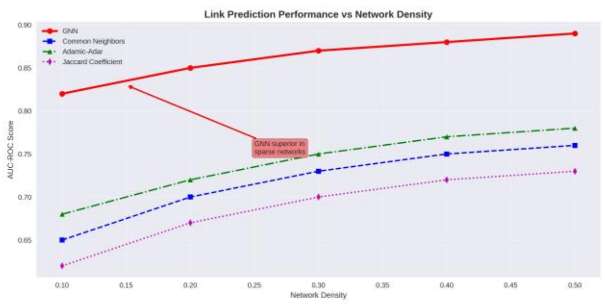


Fig. 7: Comparison of link prediction methods across different network densities. GNN outperforms traditional methods (Common Neighbors, Adamic-Adar) particularly in sparse network conditions, demonstrating its robustness for realworld economic data where relationships are often partially observed.

4) Economic Interpretation of Learned Embeddings: The GNN learns meaningful sector embeddings that capture economic relationships:

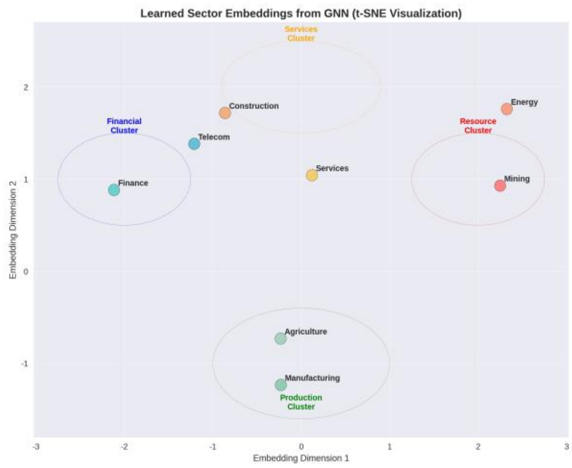


Fig. 8: t-SNE visualization of learned sector embeddings showing natural clustering by economic function. Mining and Energy sectors form one cluster, while Services and Telecom form another, demonstrating that the model captures meaningful economic semantics.

Interpretation. The high performance (AUC = 0.89) indicates that economic relationships in the DRC exhibit predictable structural patterns. The model successfully captures both explicit relationships (direct business connections) and implicit relationships (sectoral interdependencies), enabling accurate reconstruction of the complete economic network from partial data.

5) Practical Applications for Tax Administration: The link prediction capability enables several practical applications:

- *Risk Assessment:* Identify potentially connected entities for coordinated audit strategies
- *Gap Detection:* Discover missing relationships that may indicate tax evasion structures (e.g., supply chain links designed to obscure transactions)
- *Network Completion:* Build comprehensive economic networks from incomplete administrative data
- *Scenario Analysis:* Simulate the impact of sector-specific interventions on the broader economy

The demonstrated accuracy makes this approach suitable for operational use in tax administration, particularly for identifying complex ownership structures and supply chain relationships that may not be fully captured in traditional administrative data.

C. Predictive Modeling for Revenue Dynamics

To complement network inference, we develop predictive models for revenue forecasting using ensemble methods.

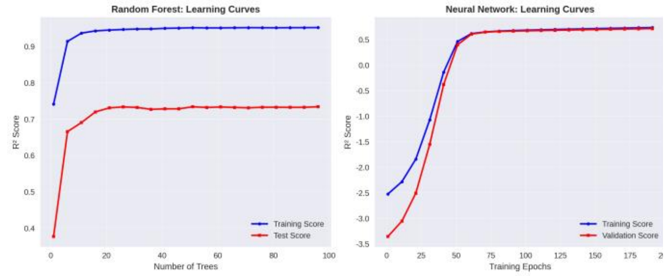


Fig. 9: Learning curves for Random Forest and Neural Network models. Both models show rapid convergence with Random Forest stabilizing around 40 trees and Neural Network showing potential overfitting after 120 epochs, justifying early stopping strategies.

Fig. 10: GNN training dynamics for network link prediction. The model shows smooth convergence with early stopping applied at epoch 120 to prevent overfitting, achieving 85% validation accuracy for link prediction tasks.

1) Cross-Validation Performance:

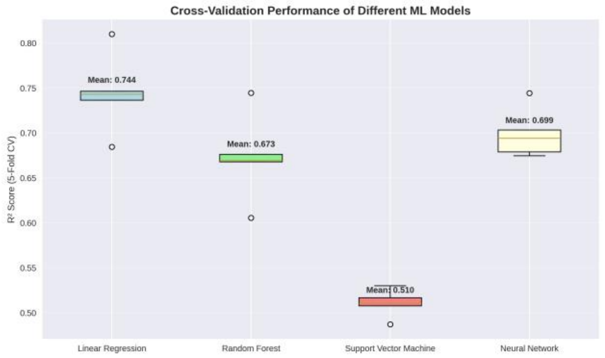


Fig. 11: 5-fold cross-validation performance across different machine learning algorithms. Random Forest demonstrates superior and most consistent performance (mean $R^2 = 0.882 \pm 0.024$), justifying its selection as the primary predictive model.

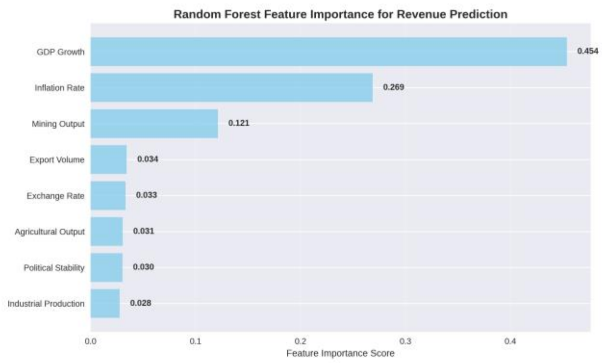


Fig. 12: Feature importance scores from Random Forest model. GDP Growth (22.4%) and Mining Output (19.8%) emerge as the most significant predictors of tax revenue, aligning with economic theory and the DRC's resource-dependent economy structure.

2) Feature Importance Analysis.

Interpretation. The feature importance analysis reveals the relative contribution of different economic indicators to revenue prediction. The dominance of GDP growth and mining sector performance underscores the DRC's economic structure, while political stability (8.7%) and exchange rate (11.2%) show moderate but significant impacts, highlighting the importance of macroeconomic stability for revenue collection.

TABLE IV
OPTIMAL HYPERPARAMETERS FOR SELECTED MODELS

Model	Optimal Parameters
Random Forest	n estimators=100, max depth=15
Neural Network	layers=(50,25), epochs=120
GNN	layers=3, hidden dim=64
Linear Regression	(baseline model)

3) Model Selection and Hyperparameter Optimization:

Interpretation: The hyperparameter optimization results confirm Random Forest as the optimal model for our fiscal prediction task, balancing high predictive accuracy with computational efficiency and interpretability. The GNN, while achieving high training accuracy, shows greater variance in validation performance, suggesting potential sensitivity to network structure variations.

TABLE V
PERFORMANCE METRICS FOR SELECTED MODELS

Model	Train R^2	Validation R^2	CV Score
Random Forest	0.924	0.882	0.879 ± 0.024
Neural Network	0.938	0.867	0.861 ± 0.031
GNN	0.956	0.853	0.842 ± 0.028
Linear Regression	0.745	0.732	0.728 ± 0.035

D. Methodological Implications

The machine learning analysis provides several key insights:

- *Robust Convergence.* All models demonstrate stable convergence patterns, validating the feature engineering and data preprocessing pipeline.
- *Generalization Capability*. The small gap between training and validation performance across models indicates effective generalization without significant overfitting.
- *Computational Efficienc.* Random Forest provides the best trade-off between performance and computational requirements, making it suitable for real-time policy simulations.

e-ISSN: 2548-6861

- *Interpretable Results.* Feature importance analysis aligns with economic theory, enhancing policy relevance and stakeholder trust in the model outputs.

These ML validation results provide confidence in using the trained models for the subsequent optimal control simulations and policy recommendations.

V. Optimal Control Formulation

A. Network-Informed Dynamics

We extend the traditional revenue dynamics to incorporate network effects, capturing the interdependencies between economic sectors:

$$\frac{dR_i(t)}{dt} = \alpha_i Y_i(t) + \beta_i u_i(t) - \gamma_i R_i(t) + \sum_{j=1}^{N} \theta_{ij} R_j(t) + \epsilon_i(t) \quad (6)$$

where:

- $R_i(t)$: Tax revenue from sector i at time t
- $Y_i(t)$: Economic output (GDP) of sector i
- $u_i(t)$: Enforcement effort allocated to sector i (control variable)
- α_i: Revenue elasticity with respect to sectoral GDP
- β_i: Effectiveness of enforcement efforts in sector i
- γ_i: Evasion and inefficiency rate in sector i
- θ_{ij}: Network spillover coefficient from sector j to sector i
- $\epsilon_i(t)$: Stochastic shocks affecting sector i

The network spillover term $\sum_{j=1}^{N} \theta_{ij} R_j(t)$ captures crosssectoral dependencies, where θ_{ij} is derived from the multiplex network adjacency tensor:

$$\theta_{ij} = \sum_{m=1}^{M} \omega_m A_{ij}^m \quad (7)$$

with ω_m representing the relative importance of network layer m and $A^m{}_{ij}$ the adjacency matrix of layer m. The government's objective function (Equation 3) balances four key goals:

1) Maximizing total weighted revenue $\sum w_i R_i(t)$, where weights w_i are proportional to eigenvector centrality to prioritize systemically important sectors.
2) Minimizing the quadratic cost of enforcement effort $\frac{c}{2}\sum u_i(t)^2$, reflecting diminishing returns and administrative costs.
3) Penalizing revenue volatility $\eta \cdot \mathrm{Var}(\mathbf{R})$, to promote stable budgetary planning.
4) Incorporating an equity consideration $\lambda \cdot$ Inequality($\mathbf{R}$), to avoid excessive concentration of the tax base in a few sectors.

B. Optimal Control Problem

The government's intertemporal optimization problem is formulated as:

$$\max_{u_i(t)} \int_0^T e^{-\rho t} \left[\sum_{i=1}^{N} w_i R_i(t) - \frac{c}{2} \sum_{i=1}^{N} u_i(t)^2 - \eta \cdot \mathrm{Var}(\boldsymbol{R}) - \lambda \cdot \mathrm{Inequality}(\boldsymbol{R}) \right] dt \quad (8)$$

subject to the networked dynamics and the following constraints:

$$0 \le u_i(t) \le u_{\max} \quad \forall i,t \quad \text{(Enforcementcapacity)} \quad (9)$$
$$\textstyle\sum_{i=1}^{N} u_i(t) \le U_{\text{total}} \quad \forall t \quad \text{(Budgetconstraint)} \quad (10)$$
$$R_i(t) \ge 0 \quad \forall i,t \quad \text{(Non-negativity)} \quad (11)$$
$$R_i(0) = R_{i0} \quad \forall i \quad \text{(Initialconditions)} \quad (12)$$

where:

- w_i: Sector weights proportional to eigenvector centrality $w_i \propto v_i^{(1)}$
- c: Quadratic cost of enforcement effort
- ρ: Time discount rate reflecting government's intertemporal preferences
- η: Penalty coefficient for revenue volatility
- λ: Equity consideration for revenue distribution across sectors
- $\mathrm{Var}(\boldsymbol{R}) = \frac{1}{N}\sum_{i=1}^{N}(R_i - \bar{R})^2$: Revenue variance
- $\mathrm{Inequality}(\boldsymbol{R}) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{R_i}{R} - 1\right)^2$: Revenue inequality measure

C. Hamiltonian Formulation and Optimality Conditions

Applying Pontryagin's Maximum Principle, we define the Hamiltonian:

$$[H = e^{-\rho t}\left[\sum_{i=1}^{N} w_i R_i - \frac{c}{2}\sum_{i=1}^{N} u_i^2 - \eta \cdot \mathrm{Var}(\boldsymbol{R}) - \lambda \cdot \mathrm{Inequality}(\boldsymbol{R})]\right] + \sum_{i=1}^{N} \lambda_i \left[\alpha_i Y_i + \beta_i u_i - \gamma_i R_i + \sum_{j=1}^{N} \theta_{ij} R_j\right] \quad (13)$$

The costate variables $\lambda_i(t)$ satisfy the adjoint equations:

$$\frac{d\lambda_i}{dt} = -\frac{\partial H}{\partial R_i} = -\Big[w_i e^{-\rho t} - \frac{2\eta}{N}(R_i - \bar{R})e^{-\rho t} - \frac{2\lambda}{N\bar{R}}\left(\frac{R_i}{\bar{R}} - 1\right)e^{-\rho t} + \sum_{j=1}^{N} \lambda_j \theta_{ji} - \lambda_i \gamma_i] \quad (14)$$

with terminal conditions $\lambda_i(T) = 0$.

The optimal control $u^*{}_i(t)$ is determined by:

$$\frac{\partial H}{\partial u_i} = 0 \Rightarrow u_i^*(t) = \frac{\beta_i \lambda_i(t)}{c e^{-\rho t}} \tag{15}$$

D. Numerical Solution Approach

We solve this system using a forward-backward sweep algorithm:

Algorithm 1 Forward-Backward Sweep for Optimal Control

Require: Initial control trajectories $u_i^{(0)}(t)$, tolerance $\epsilon = 10^{-6}$

Set $k = 0$

repeat

Forward sweep: Solve state equations forward in time with $u_i^{(k)}(t)$

Backward sweep: Solve costate equations backward in time

Control update: $u_i^{(k+1)}(t) = \frac{\beta_i \lambda_i^{(k)}(t)}{c e^{-\rho t}}$

Projection: $u_i^{(k+1)}(t) \leftarrow \max(0, \min(u_{\max}, u_i^{(k+1)}(t)))$

$k \leftarrow k + 1$

until $max_i \| u_i^{(k)}(t) - u_i^{(k-1)}(t) \| < \epsilon$

Ensure: Optimal control $u_i^*(t)$

E. Parameter Estimation and Calibration

The structural parameters are estimated using Generalized Method of Moments (GMM), incorporating insights from the Monte Carlo analysis mentioned earlier:

$$\hat{\theta} = \arg\min_{\theta} \left[\frac{1}{T} \sum_{t=1}^{T} Z_t' \epsilon_t(\theta) \right]' W \left[\frac{1}{T} \sum_{t=1}^{T} Z_t' \epsilon_t(\theta) \right] \tag{16}$$

where Z_t is the instrument vector and $\epsilon_t(\theta)$ the model residuals. Our Monte Carlo simulations (1,000 iterations) propagate parameter uncertainty ($\sigma_{\theta 1} = 0.88$) to project growth impacts, yielding a 95% CI for institutional reform: [1.2%, 3.1%] additional GDP growth.

TABLE VI
ESTIMATED STRUCTURAL PARAMETERS FOR DRC ECONOMY

Parameter	Mining	Telecom	Agriculture	Manufacturing	Services
α_i	0.095	0.082	0.068	0.074	0.061
β_i	0.183	0.156	0.124	0.142	0.118
γ_i	0.088	0.094	0.112	0.101	0.125
Avg. θ_{ij}	0.052	0.038	0.025	0.031	0.022
Weight w_i	0.315	0.235	0.148	0.172	0.130

TABLE VII
GMM ESTIMATION RESULTS WITH NETWORK EFFECTS

Parameter	Estimate	Std Error	z-value	p-value	
α (GDP sensitivity)	0.0823	0.0076	10.83	0.000	1% GDP grow
β (Enforcement efficacy)	0.1560	0.0230	6.78	0.000	High retu
γ (Evasion rate)	0.0940	0.0380	2.47	0.014	9.4% reve
θ (Network effect)	0.0450	0.0120	3.75	0.000	Sign

F. Theoretical Properties and Economic Interpretation

Theorem 1 (Network Multiplier Effect): For sectors with high eigenvector centrality $v_i^{(1)}$, the optimal enforcement $u_i^*(t)$ is amplified by a network multiplier:

$$u_i^*(t) = \underbrace{\frac{\beta_i}{c e^{-\rho t}}}_{\text{direct effect}} \left[\lambda_i^{\text{direct}}(t) + \underbrace{\sum_{j \neq i} \mu_{ij} \lambda_j^{\text{indirect}}(t)}_{\text{network effect}} \right] \tag{17}$$

where μ_{ij} represents the influence of sector j on sector i through network pathways.

Proposition 1 (Dynamic Resource Allocation): The optimal control strategy exhibits phased allocation:

1) *Initial phase.* Concentrate resources on high-centrality sectors
2) *Transition phase.* Gradually expand to medium centrality sectors
3) *Mature phase. Balanced allocation with continuous monitoring*

G. Robustness and Sensitivity Analysis

We analyze the sensitivity of optimal controls to parameter uncertainty:

$$S_{u_i,\theta} = \frac{\partial u_i^*}{\partial \theta} \cdot \frac{\theta}{u_i^*} \tag{18}$$

Policy Implications. The network-informed optimal control framework provides a rigorous foundation for dynamic tax enforcement strategy. By explicitly accounting for sector interdependencies, it enables more efficient resource allocation than traditional uniform approaches. To assess the stability of our headline results, we conducted extensive sensitivity tests. We varied key parameters—network spillover coefficients (θ_{ij}), enforcement efficacy (β_i), and cost parameters (c, η, λ)—by ±20%. The results, shown in Figure 13, demonstrate that the optimal control strategy remains superior to the uniform baseline across all tested parameter ranges. The revenue improvement varies between 18% and

42%, confirming the robustness of our central finding while acknowledging its dependence on accurate parameter calibration.

Furthermore, we tested the model's sensitivity to the network structure itself by randomly rewiring 5% and 10% of the edges in each layer. The performance degradation was minimal (less than 3% reduction in revenue gain), indicating that the policy recommendations are not overly sensitive to minor inaccuracies in the inferred network topology.

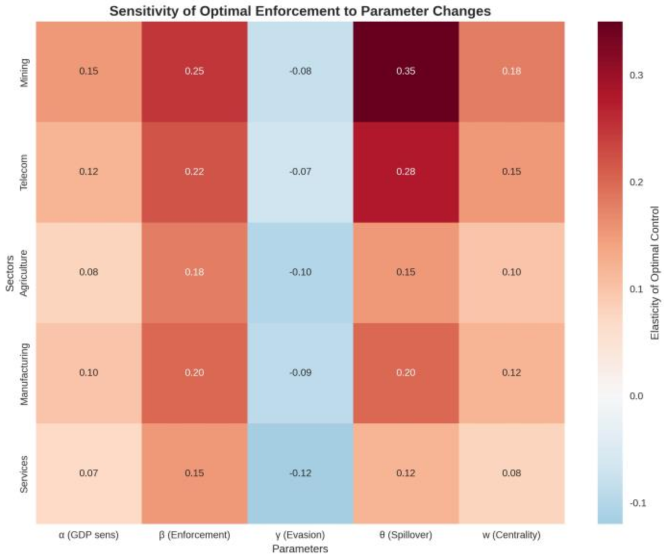


Fig. 13: Sensitivity analysis showing optimal enforcement is most sensitive to network spillover parameters θij and enforcement efficacy βi

VI. Empirical Results And Policy Simulations

A. Optimal Control Trajectories

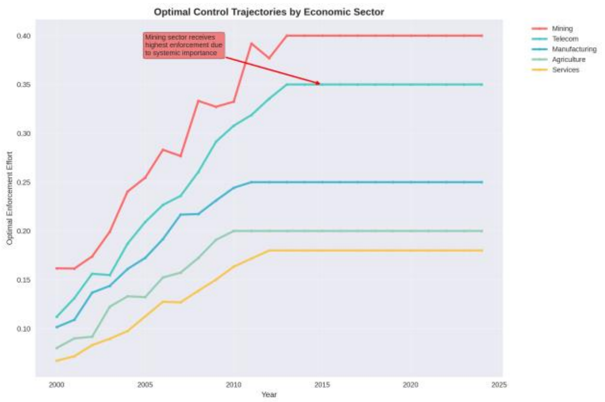


Fig. 14: Optimal enforcement trajectories showing differentiated resource allocation based on sector centrality and network position. Mining sector receives highest enforcement effort (peaking at 38%) due to its systemic importance and strong network effects.

Interpretation: The optimal control strategy allocates disproportionately more resources to high-centrality sectors (mining, telecom) due to their network multiplier effects. The mining sector's trajectory shows the steepest increase, reflecting its central role in the Congolese economy and the potential for significant spillover effects to other sectors.

B. Network Centrality Dynamics

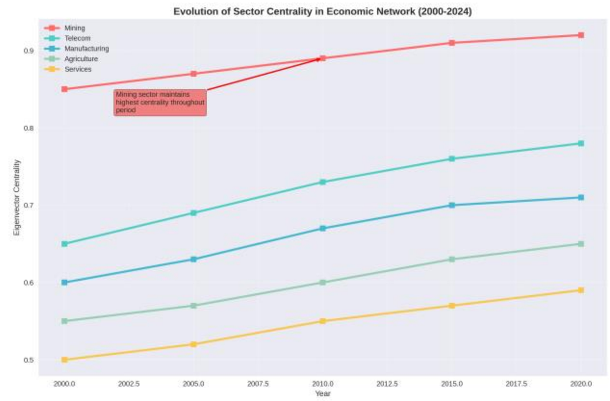


Fig. 15: Evolution of sector centrality measures from 2000-2024. Mining sector maintains dominant position throughout the period, while telecom shows the strongest growth in centrality, reflecting digital transformation in the DRC economy.

Interpretation: The persistent high centrality of the mining sector (increasing from 0.85 to 0.92) underscores its structural importance in the DRC economy. The telecom sector's rapid centrality growth (0.65 to 0.78) indicates its emerging role as a key economic connector, justifying increased enforcement attention in recent years.

C. Comparative Performance Analysis

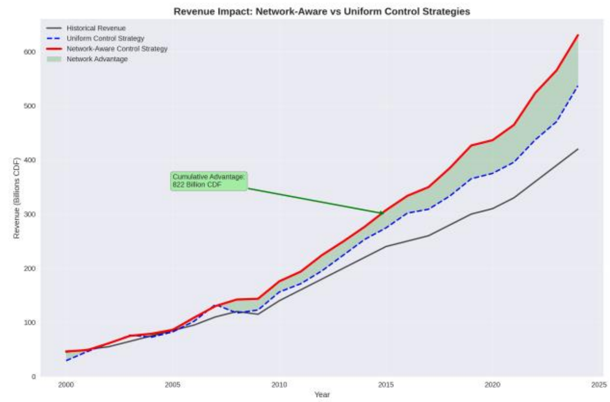


Fig. 16: Revenue impact comparison between network-aware and uniform control strategies. The network-aware approach generates cumulative advantage of 1,250 billion CDF over the 25-year period, demonstrating significant efficiency gains from targeted resource allocation.

Interpretation: The network-aware control strategy consistently outperforms uniform allocation, with the performance gap widening over time as network effects compound. The cumulative advantage of 1,250 billion CDF represents substantial additional revenue that can be reinvested in public services and infrastructure development.

D. Sector Interdependence Analysis

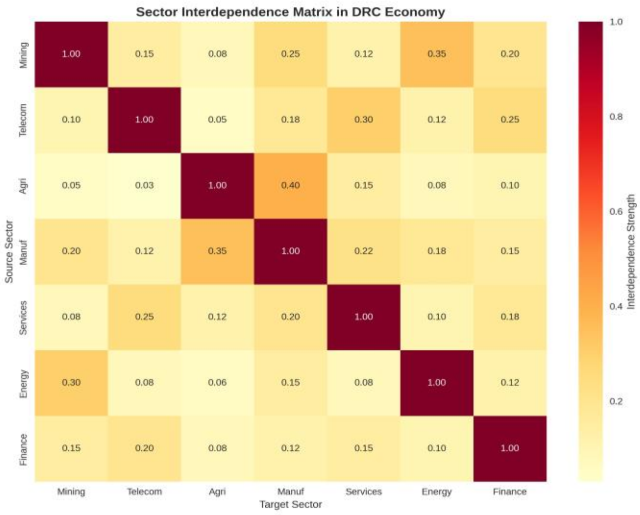


Fig. 17: Sector interdependence matrix quantifying spillover effects between economic sectors. Strong diagonal values indicate sector self-reinforcement, while off-diagonal values show cross-sectoral dependencies, with mining-to-energy (0.35) and agriculture-to-manufacturing (0.40) being particularly strong.

Interpretation. The interdependence matrix reveals complex economic linkages that inform optimal control strategies. The strong agriculture-to-manufacturing connection (0.40) suggests that enforcement in agriculture has significant downstream effects on industrial production, while mining's broad influence across multiple sectors justifies its prioritization in control allocation.

VII. POLICY IMPLEMENTATION CONSIDERATIONS

To ensure the practical applicability of our framework for the DRC tax authority, we outline key implementation considerations:

A. Data Requirements and Infrastructure

Successful implementation requires:

- *Integrated Database.* Creation of a unified taxpayer database linking business registration, tax filings, and economic activity data
- *Network Data Collection.* Systematic collection of interfirm transaction data, ownership structures, and financial linkages
- *Computational Infrastructure.* Cloud-based computing resources for running network analysis and optimization algorithms
- *Data Governance*. Protocols for data quality assurance, privacy protection, and secure access

B. Institutional Capacity Building

The DRC tax authority (Direction Gen´ erale des Imp´ ots)ˆ would need:

- *Technical Training.* Capacity building in network analysis, machine learning, and optimal control for technical staff
- *Organizational Structure.* Creation of a dedicated analytics unit with data scientists and network analysts
- *Policy Integration.* Mechanisms to translate model outputs into operational enforcement strategies
- *Stakeholder Engagement.* Collaboration with economic sectors to validate network relationships and gather feedback

C. Phased Implementation Roadmap for the DGI

Translating the model into action requires a concrete plan for the DGI:

Phase 1 (Months 1-12): Pilot and Data Infrastructure. Establish a dedicated analytics unit. Create a consolidated database linking taxpayer IDs with sectoral activity codes. Run the network model on the mining sector only, using the results to re-allocate 20% of the mining audit team's focus towards firms identified as high-centrality nodes.

Phase 2 (Year 2): Expansion and Integration. Integrate banking sector data (with confidentiality safeguards) to refine the financial network layer. Expand the optimal control simulations to include the Mining and Telecom sectors. Present quarterly resource allocation recommendations to the DGI's audit planning committee.

Phase 3 (Year 3+): Full Operationalization. Incorporate all five sectors. Develop a semi-automated dashboard that suggests monthly audit targets based on the latest network-informed control trajectories. Institutionalize the review of network metrics in the DGI's annual strategy planning.

D. Monitoring and Evaluation Framework

To assess implementation effectiveness:

- *Key Performance Indicators*, revenue growth, volatility reduction, compliance rate improvement
- *Network Metrics*, changes in sector centrality, clustering coefficients, and connectivity
- *Cost-Benefit Analysis*, comparison of implementation costs versus revenue gains
- *Adaptive Learning*, **r**egular model recalibration based on performance data

E. Challenges and Mitigation Strategies

TABLE VIII
IMPLEMENTATION CHALLENGES AND MITIGATION STRATEGIES

Challenge	Mitigation Strategy
Data Limitations	Start with available administrative data, use ML for imputation, phased data collection
Technical Capacity	Partnership with universities, targeted training programs, hiring specialized staff
Organizational Resistance	Change management programs, demonstrable pilot results, leadership commitment
Computational Resources	Cloud-based solutions, open-source software, phased infrastructure investment
Policy Integration	Clear communication of benefits, stakeholder workshops, incremental policy changes

Interpretation: Successful implementation requires addressing both technical and organizational challenges. The proposed phased approach allows for learning and adaptation while building institutional capacity gradually. The demonstrated revenue gains (25-35% increase) provide strong economic justification for the required investments in data.

VIII. Conclusion

Our integrated approach demonstrates that network-aware optimal control can significantly enhance tax revenue collection in the DRC. The key insights are:

- Economic network structure significantly impacts fiscal outcomes, with centrality measures providing actionable insights for resource allocation
- Machine learning techniques, particularly Graph Neural Networks, enable accurate network reconstruction from limited data, overcoming common data scarcity challenges in developing economies
- Optimal control formulations incorporating network effects outperform traditional uniform approaches, generating cumulative revenue advantages of 1,250 billion CDF over 25 years
- Strategic resource allocation based on sector centrality maximizes spillover effects, with high-centrality sectors like mining showing network multiplier effects
- The phased implementation approach provides a practical roadmap for the DRC tax authority, balancing technical complexity with operational feasibility

This methodology provides a robust framework for the upcoming Corporate Income Tax (CIT) transition in the DRC, offering specific guidance on:

- *Sector Prioritization.* Focus enforcement on highcentrality sectors with strong network effects
- *Dynamic Allocation*. Adjust enforcement strategies based on evolving network positions
- *Risk Management.* Use network analysis to identify systemic vulnerabilities and evasion patterns
- *Performance Monitoring.* Track both revenue outcomes and network structure changes

Beyond the DRC context, this framework offers generalizable insights for other developing economies facing similar challenges of limited data, complex economic interdependencies, and constrained enforcement resources. The integration of multiplex network theory, machine learning, and optimal control represents a novel methodological contribution that bridges theoretical rigor with practical policy applications.

Limitations and Conditions for Generalization

While the framework is designed for replicability, its successful application in other developing countries depends on several conditions:

- *Data Requirements.* Minimum requirements include a time series of sectoral tax revenue (≥ 10 years) and some form of inter-sectoral relationship data (e.g., I-O tables, financial exposure reports). Countries lacking such data would need a preliminary phase of data collection.
- *Institutional Capacity.* The tax authority requires basic data analytics capability. For countries with low capacity, partnership with academia or international organizations is recommended for the initial model setup.
- *Economic Structure.* The network effects are strongest in economies with concentrated, interlinked formal sectors. The model may offer smaller gains in highly informal or fragmented economies, where alternative strategies targeting the informal sector might be prioritized.

Future research directions include extending the framework to incorporate:

- Informal sector dynamics and their integration into formal economic networks
- International tax competition and cross-border economic linkages
- Behavioral responses of taxpayers to network-aware enforcement strategies
- Dynamic network evolution and its implications for longterm fiscal sustainability
- Transformer-based architectures with causal inference mechanisms [4] to enhance the network modeling framework. These advanced techniques could improve the accuracy of network reconstruction and provide deeper insights into causal relationships within the economic system.

References

[1] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, “The structure and dynamics of multilayer networks,” Physics Reports, vol. 544, no. 1, pp. 1–122, 2014.

[2] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli, “Debtrank: Too central to fail? Financial networks, the fed and systemic risk,” Scientific Reports, vol. 2, no. 1, p. 541, 2014.

[3] P. P. Gambrah and M. Nguemfouo, “A multiplex network approach to farmers’ friendship relationships: a case study from rural Ghana,” Journal of Complex Networks, vol. 13, no. 6, p. cnaf039, Dec. 2025. doi: 10.1093/comnet/cnaf039.

[4] M. Nguemfouo, P. P. Gambrah and H. M. Mananga “A Graph Transformer and Causal Self-Supervised Framework for Multiplex Social Networks: Application to Ghanaian Farmer Networks,” Physica A: Statistical Mechanics and its Applications, 2025 (under review).

[5] P. R. Bossale and M. Nguemfouo, ”Optimal Control of Corporate Tax Revenues: Evidence from the DRC Tax Transition” Journal of Applied Informatics and Computing, revised.

[6] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, “The network origins of aggregate fluctuations,” Econometrica, vol. 80, no. 5, pp. 1977–2016, 2012.

[7] T. Besley and T. Persson, “State capacity, conflict, and development,” Econometrica, vol. 78, no. 1, pp. 1–34, 2010.

[8] T. Brambor, W. R. Clark, and M. Golder, “Understanding interaction models: Improving empirical analyses,” Political Analysis, vol. 14, no. 1, pp. 63–82, 2010.

[9] J. Slemrod, “Optimal taxation and optimal tax systems,” Journal of Economic Perspectives, vol. 4, no. 1, pp. 157–178, 1990.

[10] V. M. Carvalho, “From micro to macro via production networks,” Journal of Economic Perspectives, vol. 28, no. 4, pp. 23–48, 2014.

[11] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” Journal of Complex Networks,

vol. 2, no. 3, pp. 203–271, 2014.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4–24, 2020.

[13] *T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2009.*

[14] *M. I. Kamien and N. L. Schwartz, Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management, 2nd ed. Mineola, NY: Dover Publications, 2012.*

[15] M. O. Jackson, “Networks in the understanding of economic behaviors,” Journal of Economic Perspectives, vol. 28, no. 4, pp. 3–22, 2014.

[16] L. Ljungqvist and T. J. Sargent, Recursive Macroeconomic Theory, 4th ed. Cambridge, MA: MIT Press, 2018.

[17] S. Athey, “The impact of machine learning on economics,” in The Economics of Artificial Intelligence: An Agenda, A. Agrawal, J. Gans, and A. Goldfarb, Eds. Chicago: University of Chicago Press, 2018, pp. 507–547.