

Optimal Allocation Model for Social Capital Resources in Community Health Management Based on Multi-Agent Deep Reinforcement Learning

Li Jin

¹Shanghai Xinnasongshan Biopharmaceutical Technology Co., Ltd, Qingdao 266100, Shandong, China

Corresponding author: Li Jin (e-mail: MA2024shannon@163.com)

ABSTRACT In community health management resource allocation, the difficulty in quantifying social capital and the inefficiency of multi-agent collaboration lead to imbalanced allocation. This problem, mirrored in the textile industry's complex supply chain where poor coordination and unquantifiable supplier relationships cause imbalanced raw material and production allocation, is addressed by this paper. We propose a model integrating dynamic social network embedding with hierarchical multi-agent deep reinforcement learning (MADRL). A graph neural network extracts node embeddings from time-series interaction data to achieve a quantitative representation of social capital. A hierarchical multi-agent architecture is constructed using heterogeneous agents responsible for demand response and resource scheduling, respectively. A joint reward function integrates health improvement, resource utilization, and network fairness, while a counterfactual baseline mechanism allocates local credit to improve policy learning accuracy. Furthermore, a parameter-sharing MAPPO (Multi-Agent Proximal Policy Optimization) algorithm with entropy regularization is employed under a centralized training and distributed execution framework to achieve stable collaborative decisionmaking. Beyond community governance, the proposed dynamic graph representation and multi-agent optimization strategy provides a computational framework for distributed information coordination and adaptive resource scheduling in intelligent electromagnetic sensing and communication environments, where reliable network interaction and efficient information propagation are critical. Experimental results demonstrate that the proposed model achieves an average Gini coefficient of 0.18–0.29 for normalized healthcare demand satisfaction and a social capital activation rate of $79.3\% \pm 3.6\%$, significantly improving allocation fairness and utilization efficiency. By quantifying social capital and enhancing multi-agent collaboration, this study provides an effective framework for alleviating distribution imbalance while offering methodological insights for intelligent resource management in complex networked engineering systems.

INDEX TERMS community health management, multi-agent deep reinforcement learning, dynamic network embedding, intelligent electromagnetic information processing, distributed communication networks

I. INTRODUCTION

COMMUNITY health management is shifting from passive service to proactive intervention, and its effectiveness increasingly depends on the ability to integrate non-institutional resources [1,2]. A similar shift is occurring in industrial management; for example, in the textile industry, successful supply chain and production management increasingly relies on the ability to actively integrate non-institutional resources, such as flexible, non-contractual, or informal production capacity and knowledge networks [3,4]. Especially in long-term governance scenarios such as chronic disease prevention and control and mental health support [5,6], the structural advantages of social relation-

ships can effectively compensate for response delays and coverage gaps in the formal service system [7,8], and drive health governance towards a more sustainable and resilient model.

Current resource allocation mechanisms in community health management mainly focus on explicit indicators such as population density, morbidity, and geographical accessibility, lacking systematic modeling of interpersonal interaction structures and their evolution [9,10]. Early studies attempted to quantify social capital through questionnaires and social network analysis, using topological indicators such as centrality and clustering coefficients to assess an individual's influence within the network, thereby guiding

the priority allocation of resources [11,12]. While these methods can provide structural insights, they rely on manually collected data, are outdated, and fail to capture subtle changes in everyday interactions [13,14]. Subsequently, some scholars applied fixed graph embedding techniques to social relationship analysis or graph representation learning in the medical field, mapping social relationships to low-dimensional vectors and improving representational power [15,16]. However, these approaches still assume a constant network structure, making it difficult to adapt to the fluidity and contextual dependence of relationships in real-world scenarios [17,18]. Furthermore, this homogeneous subject design can lead to role confusion and fail to reflect the essential differences in functional responsibilities between individuals and institutions [19,20]. Overall, existing models have not yet formed an effective closed loop between dynamic modeling of social capital, expression of subject heterogeneity, and joint decision-making mechanisms, which limits the adaptability and robustness of resource allocation in complex community ecosystems.

This research aims to construct a social capital-driven resource allocation model for community health management. Its core focus is on achieving computability of social capital and intelligent multi-agent collaboration.

II. CONSTRUCTION OF MULTI-AGENT COLLABORATIVE DECISION-MAKING MODEL DRIVEN BY DYNAMIC SOCIAL NETWORKS

Figure 1 illustrates the overall structure of a community health management resource allocation model based on multi-agent deep reinforcement learning. The dynamic social network module extracts social capital features from resident interaction data using a graph neural network, outputting node embeddings that serve as agent input. The hierarchical multi-agent system consists of an individual layer and an organizational layer, responsible for health response and resource scheduling, respectively. The centralized training framework uses the MAPPO algorithm to optimize sharing strategies and incorporates entropy regularization to enhance exploration capabilities. The joint reward function integrates health, resource utilization, and fairness objectives, and local credit allocation is achieved through counterfactual baselines. Ultimately, decentralized decision-making is achieved through distributed execution, forming a closed-loop optimization system from social perception to collaborative regulation.

DYNAMIC REPRESENTATION OF SOCIAL CAPITAL BASED ON TIME-SERIES BEHAVIORAL DATA

Dynamic Embedding Modeling of Social Capital Based on Time-Series Graph Neural Networks

To achieve continuous quantitative representation of social capital among community members, this study constructs a dynamic graph structure using time-series behavior logs as input. Community individuals are considered nodes in the graph, and their interaction events within specific time

windows—including communication frequency, information forwarding paths, and actual mutual assistance behaviors—serve as edge weights and existence criteria, forming a series of snapshot-like dynamic adjacency matrices. These graph sequences at discrete time segments are fed into a time-aware encoding module based on a gated graph neural network (GGNN), which updates the hidden state vector of each node cycle by cycle. GGNN is chosen primarily because it effectively captures long-term dependencies and state persistence in social relationship evolution through its gating mechanism, demonstrating superior temporal modeling capabilities compared to standard Graph Attentional Networks (GAT) when processing discrete time snapshot sequences. Specifically, let the graph structure at the t -th time step be $G_t = (V, E_t)$, where V represents the set of all resident nodes in the community, and E_t is the set of interaction edges existing at that moment. The hidden state of each node $i \in V$ at time t is denoted as $h_i^t \in \mathbb{R}^d$, and the dimension d is the size of the embedding space. The state update follows the following iterative rule:

$$h_i^t = \text{GRU} \left(h_i^{t-1}, \sum_{j \in N_i^t} W_e^t \cdot m_{ji}^t \right) \quad (1)$$

N_i^t represents the set of neighbors that have direct interactions with node i at time t ; m_{ji}^t is the message vector transmitted from node j to i , which is usually determined by the edge type and strength weighting; W_e^t is the learnable message transformation parameter matrix; GRU (Gate Recurrent Unit) represents the operation function of the gated recurrent unit. The dynamically generated node hidden state h_i^t represents the individual's social capital at time t . This update mechanism allows the model to automatically identify high-frequency interaction groups, key information intermediaries, and potential trust hubs, thereby generating a dynamic embedding sequence that reflects the evolution trend of social relations. From a sociological perspective, this dynamic embedding comprehensively reflects the multi-dimensional social capital forms within individual community networks. High self-loop weights and local clustering responses correspond to "connective" social capital, reflecting trust and norms within groups. The activation of high betweenness centrality across community connections represents "bridging" social capital, facilitating crossboundary information flow. Meanwhile, long-term stable reciprocal interaction patterns are reinforced through temporal modeling, which is associated with the accumulation of "connective" social capital. Before the start of each decision cycle, the system calls the node embedding h_i^t output in the latest round as an important component of the state space of the corresponding individual agent, ensuring that the policy network can respond based on the latest social structure cognition.

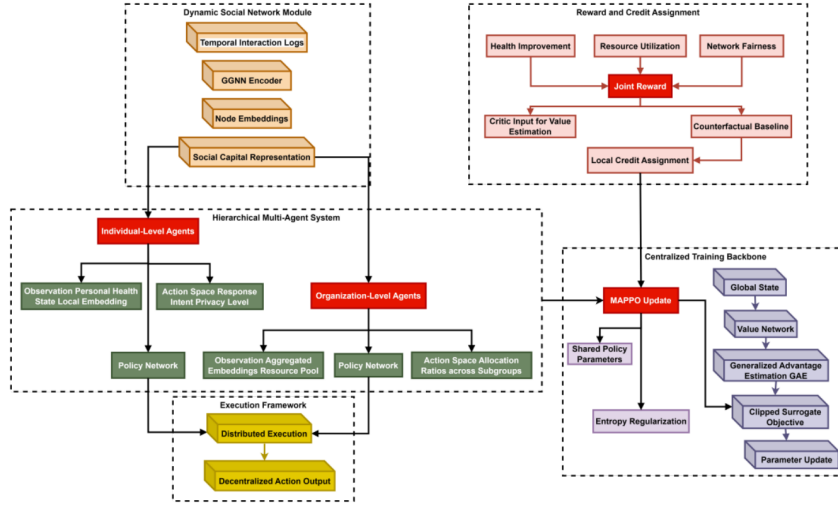


FIGURE 1. Community health management resource allocation model

EMBEDDING ENHANCEMENT MECHANISM INCORPORATING MULTI-DIMENSIONAL INTERACTION SIGNALS

To further enhance the granularity and semantic richness of social capital representation, the model applies a multi-channel input processing mechanism, distinguishing different types of social behaviors and assigning differentiated communication weights. The raw behavior logs are divided into three semantic channels: communication frequency, knowledge dissemination, and substantive support behaviors, corresponding to daily conversations, health information forwarding links, and records of material assistance or psychological support, respectively. A separate edge weight system is constructed for each behavior type, and a channelby-channel attention mechanism is used during graph convolution, enabling the model to independently determine the relative importance of each type of interaction in shaping social capital. This design avoids information masking caused by simple aggregation and particularly enhances sensitivity to non-explicit but high-value interactions (such as emotional support).

In each embedding update, the system calculates attention coefficients between channels to adjust the contribution ratios during message aggregation. For node i , the comprehensive information it receives from neighbor j at time t is defined as a weighted sum:

$$m_i^t = \sum_{k=1}^3 \alpha_k^t \left(\sum_{j \in N_i^t} A_{ji}^{(k)} \cdot W_k h_j^t \right) \quad (2)$$

$A_{ji}^{(k)}$ is the adjacency matrix element composed of the k -th type of behavior; $W_k \in \mathbb{R}^{d \times d}$ is the linear transformation matrix dedicated to the channel; α_k^t is the dynamic weight coefficient generated by the lightweight attention network, satisfying $\sum_{k=1}^3 \alpha_k^t = 1$ and $\alpha_k^t \geq 0$. This attention mechanism evaluates the activation level of each channel

in the current community context through a small-scale feedforward network.

HIERARCHICAL MULTI-AGENT SYSTEM ARCHITECTURE DESIGN

Role-based State Perception and Observation Space Construction for Heterogeneous Agents

To truly reflect the diverse functional roles of multiple agents in community health management, the system adopts a hierarchical architecture and deploys two types of functionally decoupled agents: individual-level agents and organizational-level agents. All agents share a unified global state input during the decisionmaking cycle. This state is composed of dynamic social network embeddings, resource inventory levels, regional health indicators, and environmental constraints, ensuring that all participants have a consistent understanding of the situation. Based on this, the observation space is differentially projected based on role attributes. The observation vectors of individual-level agents focus on the evolution of their own health status, changes in the intensity of neighborhood interactions, and the strength of received intervention signals. A nonlinear mapping module extracts a subset of local features relevant to their behavioral responses. Organization-level agents receive topologically aggregated group statistics, including subnetwork average embedding centrality, service coverage gap index, and resource consumption rate, to support macroscheduling decisions.

To implement a role-adaptive information screening mechanism, an observation selection function based on learnable weights is designed, enabling agents at different levels to extract semantic components from the global state that align with their responsibilities. Assuming that the global state is represented as s_t , the local observation o_i^{ind} of the i -th agent at the individual level is generated by the following transformation:

$$o_i^{\text{ind}} = \sigma(W_{\text{ind}}(s_t \odot a_i) + b_{\text{ind}}) \quad (3)$$

W_{ind} is the individual-specific projection matrix; b_{ind} is the bias term; σ represents the nonlinear activation function; $\odot$ represents the element-wise product operation; a_i is a trainable attention mask vector used to enhance the activation response of dimensions that are strongly related to the individual's health response. Similarly, the observation o_j^{org} of the j -th scheduling agent at the organizational level is independently modeled by another set of parameters, and its attention distribution is more inclined to the resource distribution imbalance and network structure vulnerability indicators. This mechanism avoids role confusion caused by homogenized perception and ensures semantic coordination between individual demand response and organizational resource allocation.

Definition of Hierarchical Action Space and Collaborative Decision Flow Control Mechanism

At the action execution level, the two types of agents are given essentially different strategy output structures to match their actual functional positioning in the resource flow chain. The action space of the individual-level agent is defined as a discrete-continuous hybrid, including the willingness to respond to health intervention requests (binary choice) and the adjustment of the degree of privacy disclosure (continuous scaling). This output directly affects its acceptance threshold for external resource calls; the action space of the organizational-level agent is expressed as a multi-dimensional continuous vector, each dimension corresponding to the distribution ratio of a specific type of resource (such as psychological counseling quota, chronic disease follow-up frequency, health education materials) among several geographical or social subgroups, which must meet the total conservation and non-negativity constraints. To coordinate the decision logic of the upper and lower levels, a parameterized decoupling action generation framework is applied. The organizational-level policy network outputs the resource allocation coefficient vector u_j , which is Softmax normalized and then tensor-scaled with the available resource pool R_t to form the actual delivery instruction:

$$r_{j,k} = R_t \cdot \frac{\exp(u_{j,k})}{\sum_{k'=1}^K \exp(u_{j,k'})} \quad (4)$$

$r_{j,k}$ represents the amount of resources allocated by the j -th organization to the k -th subgroup, and K is the total number of subgroups. This instruction is injected into the state space of the individual layer as an external excitation signal, triggering the behavioral response of the resident agent. At the same time, the response intensity summarized at the individual layer is weighted and fed back to the next cycle input of the organizational layer, forming a closed-loop control pathway.

To achieve closed-loop feedback, the comprehensive response intensity at the individual level is quantified through

outputs from their action space. Specifically, the "response willingness" (binary value) of all individual agents for resource requests is aggregated, while their "privacy disclosure level" (continuous value) is averaged to form two scalar indicators. After normalization, these two indicators are linearly combined with equal weighting to create a comprehensive response index. This index serves as one of the state features, which is then injected into the global state input at the organizational level for the next decision cycle.

JOINT REWARD FUNCTION AND LOCAL CREDIT ALLOCATION MECHANISM CONSTRUCTION

Joint Reward Construction and Dynamic Weighting Mechanism for Multi-objective Coupling

To achieve the coordinated optimization of multiple governance objectives in community health management, a composite reward signal consisting of three elements: health gain, resource utilization, and social equity, is constructed. This reward is calculated based on system-level observational data at the end of each decision cycle, providing quantitative feedback on the effectiveness of overall strategy execution. The health improvement dimension is evaluated by tracking trends in key health indicators, focusing on the long-term impact of interventions on chronic disease risk factors and mental health status. Resource utilization measures the balance between the actual consumption of deployed service resources and redundant losses, avoiding over-concentration or ineffective operation. Network fairness assessment focuses on structural differences in resource coverage, using inter-group distance metrics in the embedding space to identify potentially marginalized groups and weighted compensation for their improved access opportunities. After normalization, the three sub-goals are linearly combined with learnable weights to form the final global reward:

$$r_t = \omega_h(t) \cdot \phi_h(s_t) + \omega_u(t) \cdot \phi_u(s_t) + \omega_f(t) \cdot \phi_f(s_t) \quad (5)$$

$\phi_h(s_t)$, $\phi_u(s_t)$, and $\phi_f(s_t)$ represent the health improvement score, resource utilization score, and network fairness index under state s_t , respectively, all normalized to the interval $[0,1]$; $\omega_h(t)$, $\omega_u(t)$, and $\omega_f(t)$ are weight coefficients that are adaptively adjusted as the training progresses, satisfying $\sum_{i \in \{h,u,f\}} \omega_i(t) = 1$. The update mechanism is based on the adaptive weighting principle of return volatility, which dynamically adjusts the incentive intensity by minimizing the volatility of the historical returns of each sub-target. Specifically, the weights of each sub-objective are dynamically adjusted according to the inverse of its return volatility in the past cycle, that is, the sub-objective with more stable performance is given a higher weight, so as to indirectly achieve the minimization of historical strategy gradient variance. This design avoids the goal rigidity problem caused by fixed weights, enabling the system to automatically emphasize the most sensitive or lagging performance dimension at different stages, thereby

maintaining a dynamic balance among multiple goals. The global reward serves as the input of the centralized value network and is used to guide the overall optimization direction of the policy parameters.

Figure 2 illustrates the dynamic evolution of multi-objective joint rewards. As training iterations progress, the health gain curve first shows rapid improvement, reflecting the direct effect of individual interventions on short-term health indicators. The resource efficiency curve then steadily rises, demonstrating that the organizational layer optimizes resource utilization and redundancy control through gradual learning. The fairness index improves more slowly, indicating that the structural inertia of the social network and the coverage gaps of marginalized groups require a longer cumulative effect to improve.

LOCAL CREDIT ALLOCATION STRATEGY BASED ON COUNTERFACTUAL BASELINES

In the face of the widespread credit attribution ambiguity in multi-agent collaboration, a method for decomposing individual contributions based on counterfactual reasoning is applied, aiming to separate the true marginal effect of a specific agent's behavior from the collective benefit. The traditional shared reward mechanism is prone to policy confusion, especially when multiple agents take action at the same time, making it difficult to distinguish between key drivers and passive followers. To this end, after each round of policy sampling, the counterfactual baseline is estimated via a centralized Critic network. For the current agent, while fixing the actions of other agents, its own actions undergo expectation sampling under the policy distribution to calculate its marginal contribution. This baseline estimation is entirely centralized, eliminating the need to explicitly simulate each agent's environmental transitions; the resulting change in the value function is calculated as its independent influence estimate.

Specifically, the counterfactual advantage term of the i -th agent at time t is defined as:

$$A_i^{\text{cf}}(s_t, a_t) = Q(s_t, a_t) - \mathbb{E}_{a'_i \sim \pi_i(\cdot|o_i)} [Q(s_t, (a_{-i}, a'_i))] \quad (6)$$

$Q(s_t, a_t)$ is the state-action value estimate under the current joint action; a_{-i} represents the action vectors of the other agents except agent i ; the expectation term represents the change in the average Q value corresponding to the action of agent i while keeping the behavior of other agents unchanged. All Q values are parameterized by a shared centralized Critic network, while the counterfactual baseline is entirely estimated through this centralized Critic network, eliminating the need to explicitly simulate each agent's environmental transitions. This difference reflects the degree of disturbance to the overall performance when the agent deviates from the original policy, that is, the causal importance of its decision. Furthermore, the counterfactual advantage is

used as a substitute for the advantage in the policy gradient, and the policy update target is reconstructed:

$$L_{\text{policy}} = \mathbb{E}_t \left[\sum_i \log \pi_i(a_i^t | o_i^t) \cdot A_i^{\text{cf}}(s_t, a_t) \right] \quad (7)$$

This form effectively weakens the “riding” phenomenon and improves the learning sensitivity of high-contribution individuals. At the same time, the batch parallel evaluation of counterfactual Q -values from Critic network achieves controllable computational cost and is independent of environmental simulation. By sinking the attribution analysis of global rewards to the individual level, this mechanism significantly enhances the directional accuracy and responsibility transparency of policy updates. Especially in complex scenarios where residents' responses fluctuate violently or organizational scheduling is frequently adjusted, it can still stably identify the key decision nodes that truly drive the system towards optimal evolution.

COLLABORATIVE POLICY OPTIMIZATION PROCESS BASED ON POLICY GRADIENTS

Parameter-Sharing Policy Optimization Framework Based on Centralized Value Learning

To achieve stable convergence for multi-agent systems in high-dimensional heterogeneous decision spaces, the Multi-Agent Proximal Policy Optimization (MAPPO) algorithm within a centralized training with decentralized execution (CTDE) architecture is used to construct a collaborative and scalable learning process. During training, the policy networks of all agents are organized using a parameter-sharing mechanism: agents of the same type (such as individual-level resident agents or organizational-level scheduling agents) share the same set of policy network weights, significantly improving sample utilization and mitigating overfitting risks. While each agent generates actions based on local observations, its policy gradient updates rely on a globally observable central value function. This function, which takes the joint state and joint action as input, provides an accurate estimate of the collective long-term reward.

The value network uses a two-stream design, embedding state representations and action embeddings, respectively. A cross-attention module captures high-order interactions between states and actions. Its output is used to compute the generalized advantage estimation (GAE), which serves as the advantage benchmark for policy updates. Considering the sensitivity of policy updates in a multi-agent environment, a clipping mechanism is applied to limit the deviation of a single-step policy. Specifically, the probability ratio of the i -th agent at step t is defined as:

$$\rho_i^t(\theta) = \frac{\pi_{\theta}(a_i^t | o_i^t)}{\pi_{\theta_{\text{old}}}(a_i^t | o_i^t)} \quad (8)$$

π_{θ} represents the action probability distribution under the current policy network parameters, and $\pi_{\theta_{\text{old}}}$ is the fixed

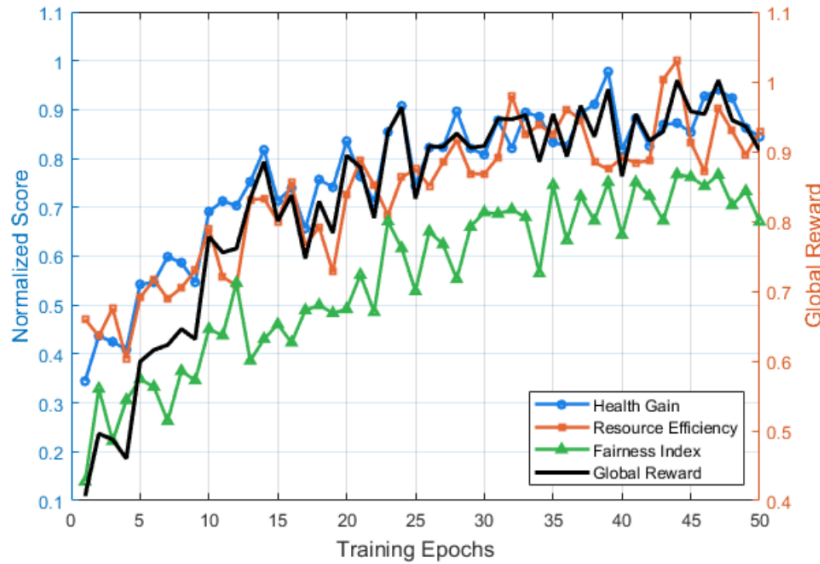


FIGURE 2. Dynamic evolution of multi-objective rewards

policy before the previous round of updates. The objective function is constructed as follows:

$$L_{\text{clip}}(\theta) = \mathbb{E}_t \left[\min \left(\rho_i^t(\theta) \hat{A}_t, \text{clip} \left(\rho_i^t(\theta), 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t \right) \right] \quad (9)$$

$\hat{A}_t$ is the advantage estimate calculated by GAE, and ϵ is the preset clipping threshold, which controls the conservatism of the policy update. This objective function ensures that the policy only accepts large changes when the benefits are clearly improved, avoiding performance fluctuations caused by value estimation bias. All agents perform gradient calculations and parameter updates synchronously in the same training batch to ensure the coordinated consistency of policy evolution. The parameter sharing mechanism combined with centralized advantage evaluation significantly enhances the generalization ability and training stability of the model in large-scale community scenarios.

Entropy Regularization-driven Exploration - Utilizing Dynamic Balance Mechanism

To prevent the strategy from converging to a local feasible solution too early in the early training stage, especially when the resource allocation pattern has not been fully explored and falls into an inert response mode, the entropy regularization term is applied to continuously stimulate the diversity of the strategy distribution.

An entropy penalty term is added to the original strategy objective function to form a comprehensive optimization objective:

$$L_{\text{total}}(\theta) = L_{\text{clip}}(\theta) + \beta \cdot \mathbb{E}_t [H(\pi_\theta(\cdot|o_i^t))] \quad (10)$$

$H(\pi_\theta(\cdot|o_i^t))$ represents the Shannon entropy of the strategy distribution under observation o_i^t , which measures the uncertainty level of action selection; β is the temperature

coefficient, which controls the weight of the exploration intensity. This coefficient is not a fixed constant, but adaptively decays with the training rounds: a higher value is given in the early stage to encourage extensive trial, and a gradually reduced value is given in the later stage to focus on the high-return strategy area. The gradient of the entropy term is directly backpropagated to the output of the last logits layer of the policy network, driving network parameter adjustments toward improved distribution uniformity.

Table 1 lists the core hyperparameter configurations for the policy optimization process of a community health management model based on multi-agent deep reinforcement learning. The entropy regularization term is used to maintain policy exploration and prevent regression into local optima. The batch length is set to balance computational efficiency with the requirements of temporal dependency modeling. All parameters are calibrated through grid search and ablation experiments, making them suitable for largescale heterogeneous agent collaboration scenarios.

TABLE 1. Core training parameters for multi-agent reinforcement learning

Parameter	Value	Description
Policy Learning Rate	3.0e-4	Step size for policy network updates
Value Learning Rate	5.0e-4	Update rate for value function optimization
Clipping Coefficient	0.2	Bounds policy change to ensure stability
GAE Lambda	0.95	Controls bias-variance trade-off in advantage estimation
Discount Factor	0.99	Weight for future reward accumulation
Entropy Regularization	0.01	Promotes exploration during training
Sequence Length per Batch	64	Number of sequential steps in each batch

Figure 3 illustrates the key dynamics of the model training

process, specifically the relationship between policy optimization and the exploration/exploitation balance. Figure 3(a) depicts the trends of Clipped Policy Loss and Entropy. As training iterations progress, Clipped Policy Loss decreases from 0.6 to near 0, indicating that the model's policy is continuously optimizing and moving toward more effective decision-making. Conversely, Entropy decreases, reflecting a gradual decrease in exploration and a more deterministic model. This change generally indicates that the model has transitioned from extensive initial exploration to more precise exploitation of the optimal policy. Figure 3(b) shows the changes in the Action Diversity Index and the β Coefficient. The Action Diversity Index shows a gradual decrease as training progresses, indicating that as the policy matures, the model's action diversity gradually decreases and tends to a stable behavior pattern. The β Coefficient also continues to decline, indicating that the model's entropy regularization gradually weakens during training, emphasizing the gradual convergence of exploration. This change reflects a typical phenomenon in reinforcement learning—the transition from high exploration to high utilization, ensuring a balance between training efficiency and effectiveness.

III. COMPARISON OF MULTI-DIMENSIONAL COLLABORATIVE ALLOCATION EFFECTIVENESS EXPERIMENTAL DATA

The experimental data is derived from a long-term health monitoring platform in three typical urban communities in a certain city, covering multi-source heterogeneous information from January 2022 to June 2024. The dataset includes residents' electronic health records, community health service center service records, daily interaction logs collected by smart terminals (such as neighborhood mutual assistance, health group exchanges, and event attendance), and public facility usage trajectories. After desensitization, a dynamic social network graph sequence is constructed. A total of 5,832 permanent residents are included, forming a collection of interaction snapshots with a sampling granularity of 24 hours, generating over 1.2×10^6 valid interaction edges. Individual attributes include age, chronic disease status, health literacy scores, and previous intervention response records. Organizational nodes include three community health stations, four social work organizations, and two volunteer teams, and their resource supply capacity and time-varying load levels are modeled in real-time.

All data have undergone complete desensitization, with direct identifiers (e.g., name, ID number, address) removed, retaining only aggregated statistical features and anonymized interaction records. Data collection adhered to the informed consent principle and was approved by the internal ethics review process of the collaborating institution. The original logs are accessible only to the research team within a secure computing environment and are not publicly disclosed.

An "effective interaction edge" is defined as: within a 24-hour window, at least two bidirectional interactions (e.g.,

message exchanges, mutual aid record matching, or joint activities) between two nodes, with content filtered by rules to exclude low-semantic noise such as system notifications or auto-forwarding. For missing interaction snapshots (e.g., unreported devices), a forward filling combined with community mean interpolation strategy is applied to ensure graph sequence continuity.

The model training/test was conducted with strict time segmentation: data from January 2022 to December 2023 was used for training, and data from January to June 2024 was used for testing. Cross-period information leakage was prohibited to ensure that the evaluation reflects true generalization ability.

ATTENTION WEIGHTS AND EMBEDDING SIMILARITY FOR MULTI-CHANNEL INTERACTION BEHAVIORS

Based on the node embedding vectors generated by the temporal graph neural network, the cosine similarity matrix is calculated between all node pairs. For multi-channel attention weight evaluation, the normalized weights of each channel at each time step are extracted from the recorded attention activation values during the forward propagation process to ensure that they reflect the relative importance of different interaction types in the dynamic network. Community structure evaluation utilizes cluster consistency analysis. Heatmap visualization of the similarity matrix identifies highly cohesive node groups, and a betweenness centrality algorithm is used to detect bridge nodes and verify their cross-community connectivity. Figure 4(a) shows the precise evolution of attention weights for three types of interaction behaviors (communication frequency, information sharing, and supportive behavior) over five time steps. Initially ($t=1$), the three types of behaviors are assigned similar weights (approximately 33% each), reflecting the model's fundamental emphasis on each type of interaction. Over time, the weight of the information sharing channel increases significantly, peaking at 50% at $t=3$, corresponding to the community context of intensive health information dissemination. At $t=4$, the weight of supportive behavior rises significantly, reaching 45%, reflecting the high attention paid to substantive mutual assistance in crisis intervention scenarios. Finally, at $t=5$, the weight of communication frequency rises to 40%, becoming the dominant channel, demonstrating the enduring role of daily communication in maintaining social capital. This dynamic evolution demonstrates that the attention mechanism can adaptively adjust the importance of different interaction types, precisely responding to evolving community contexts. Figure 4(b) shows the cosine similarity matrix of 50 nodes based on the final embedding vectors, exhibiting a clear block diagonal structure and a distinct bridge node pattern. Color represents cosine similarity. The matrix shows three distinct high-similarity communities: nodes 1-15 form the first tight-knit community (red block); nodes 16-35 form the second; nodes 36-50 form the third. Communities maintain high internal similarity, while inter-community similarity (blue area) is lower, accurately reflect-

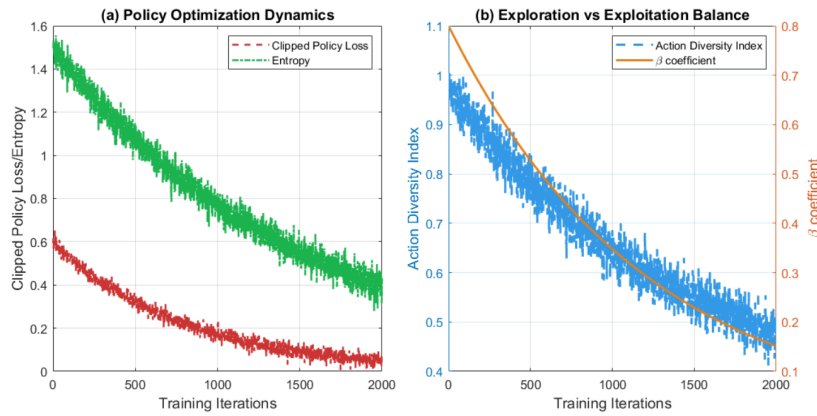


FIGURE 3. Policy optimization and exploration/exploitation balance

ing the natural divisions of social groups in real communities. Particularly noteworthy are nodes 25 and 40, acting as bridge nodes, exhibiting moderate similarity with multiple communities (transition zone, 0.4-0.6). This indicates that they lie at the intersection of different communities in the embedding space, corresponding to their real-world role as information intermediaries, facilitating cross-community information flow and resource transfer.

SPATIOTEMPORAL DISTRIBUTION OF ATTENTION WEIGHTS OF HETEROGENEOUS AGENTS

To evaluate the effectiveness of heterogeneous agent attention mechanisms, six core feature indicators are extracted based on time-series interaction data from a real-world community health management platform. Dynamic node embeddings are generated through a gated graph neural network, and a learnable attention mask mechanism is used to perform role-based projections of the global state. A parameterized softmax function is used to calculate the attention weight distribution of the two agents at each decision cycle, the weight coefficient is adjusted according to the inverse of the volatility of the historical return of each subobjective.. Finally, the temporal evolution of attention weights is visualized through heatmaps.

Figure 5 shows the dynamic changes in the attention weights of individual-level and organizational-level agents over 100 time steps, arranged in a top-down heatmap format. Color represents attention weight. Figure 5(a) shows the attention distribution of individual-level agents, demonstrating their consistent high attention to the last three features: health trends, local interaction strength, and intervention signal strength (which gradually increase over time). However, their attention weights for the first three macro-level features (embeddedness centrality, coverage gap index, and resource consumption rate) are relatively low and stable, consistent with the individual-level agent's primary role in responding to local health needs. Figure 5(b) illustrates the attention pattern of organizational-level agents, demonstrating a completely different distribution. They pay relatively little attention to individual-level features, while paying close

attention to three macro-level statistical features, particularly embeddedness centrality (whose attention weight gradually increases over time) and coverage gap index (which exhibits a cyclical pattern of attention). Their resource consumption rate also maintains a relatively stable, moderate level of attention.

RESOURCE ALLOCATION FAIRNESS

To systematically evaluate the fairness of different resource allocation mechanisms in spatially heterogeneous communities, the Gini coefficient and Theil index are used as core indicators to measure the degree of resource inequality. It is important to note that the Gini coefficient is calculated based on a normalized measure of need satisfaction (i.e. the ratio of resources allocated to estimated health needs) and uses risk weights at the subgroup level, rather than the original amount of resources, to reflect equity in health care, rather than simple quantitative equality. Experiments are conducted in three density scenarios: low (<50 people/km²), medium (50–150 people/km²), and high (>150 people/km²). Each community type generates an independent simulation environment based on the actual geographic distribution of the population. The proposed MADRL model is compared with traditional MADRL(QMIX, as a representative of traditional collaborative MADRL), random allocation, linear programming (LP), and greedy algorithms (GAs). Resource allocation strategies are executed within consecutive scheduling cycles, and resource coverage data is recorded at the end of each cycle. Through cross-period statistical analysis, the mean and standard deviation of the fairness indicators of each method under three types of density conditions are calculated to quantify the stability and adaptability of the model in different social structure environments. Figure 6 shows a comparative analysis of the fairness performance of five resource allocation methods at different community densities. Figure 6(a) uses the Gini coefficient to assess the degree of inequality in resource coverage. The results show that the proposed MADRL model maintains the lowest level of inequality in low, medium, and high density environments, with average Gini

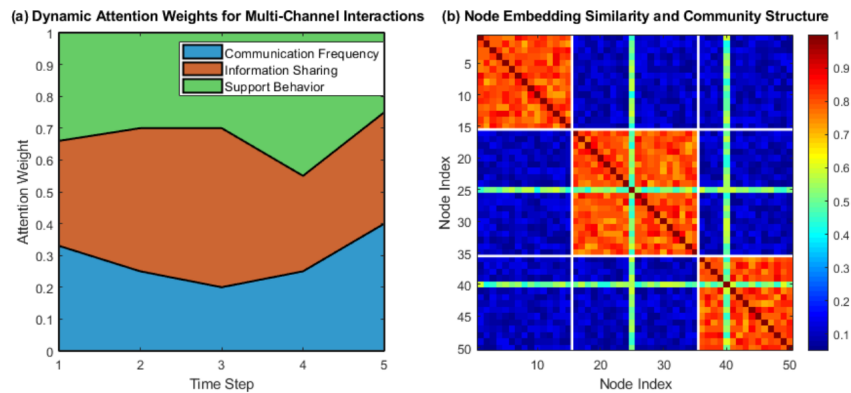


FIGURE 4. Analysis of attention weights and embedding similarity for multi-channel interaction behaviors



FIGURE 5. Comparative analysis of attention weights

coefficients ranging from 0.18 to 0.29. This demonstrates that through dynamic social network perception and hierarchical decision-making mechanisms, it effectively identifies and responds to the needs of marginalized groups, avoiding excessive concentration of resources in highly connected nodes. Traditional MADRL and greedy algorithms, lacking fine-grained modeling of social capital structure, are prone to allocation polarization. Figure 6(b) shows the trend of the Theil index as a function of spatial density, revealing the system's adaptability to structural differences. The proposed method maintains low and stable index values across all three density scenarios, with average Theil index values ranging from 0.20 to 0.31, demonstrating the robustness and generalizability of its strategy. In contrast, linear programming and random allocation fail to fully account for individual response heterogeneity and network transmission effects, resulting in a significant decline in fairness with increasing environmental complexity.

RESPONSE EFFICIENCY AND STABILITY EVALUATION

To comprehensively evaluate the response efficiency and stability of various resource allocation strategies in public health emergencies, a simulation test environment based on epidemic evolution is designed, using resource allocation delay time and the number of convergence steps for health risk control as core performance indicators. The experiment simulates the community emergency response process under different impact intensities using three initial infection conditions: mild (5%), moderate (15%), and severe (30%). Thirty independent Monte Carlo simulations are run for each scenario to eliminate random fluctuations. By calculating the time interval between resource scheduling initiation and key node coverage completion, as well as the number of decision cycles required for the health risk index to steadily decrease and enter a controllable threshold, the response speed and system stability of each method are quantified, and a multi-level comparative analysis framework is constructed.

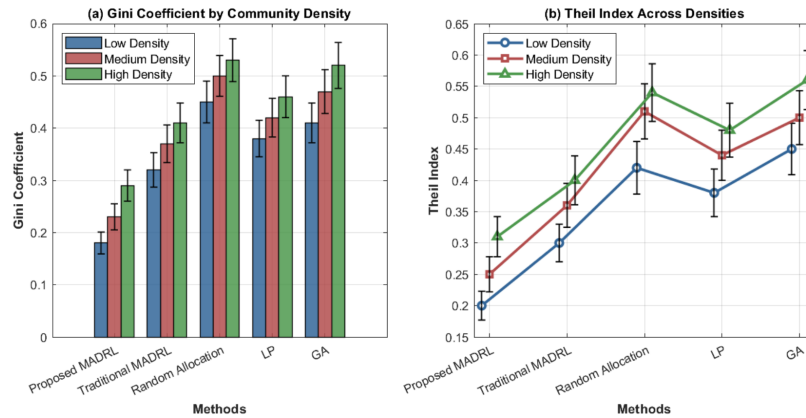


FIGURE 6. Fairness performance at different community densities

Figure 7 compares the response efficiency and system stability of different resource allocation strategies under sudden epidemic scenarios. Figure 7(a) shows the resource allocation delays of five methods under mild, moderate, and severe epidemic impacts. The proposed MADRL model exhibits shorter response delays in all scenarios, maintaining particularly high scheduling agility under severe conditions, with a resource allocation delay of 9.1 ± 1.2 decision cycles. Its advantage stems from the synergy between dynamic social network perception and a hierarchical decision-making mechanism, enabling rapid and targeted deployment of critical resources along trust paths. Figure 7(b) shows the convergence trend of health risk control. The proposed method achieves a convergence of 21 ± 2.2 decision cycles under severe epidemic impact, demonstrating greater robustness. In contrast, traditional MADRL suffers from a sluggish response due to its lack of a credit allocation mechanism. Linear programming and greedy algorithms are limited by static modeling and struggle to adapt to dynamic evolution. Random allocation exhibits the greatest volatility, validating the necessity of intelligent collaborative decision-making in emergency health management.

SOCIAL CAPITAL UTILIZATION EFFICIENCY

To quantify the efficiency of resource allocation strategies in utilizing social capital, a dual evaluation system for social capital utilization is constructed. The social capital activation rate is defined as the ratio of the number of support relationship chains actually triggered during the intervention process to the total number of potentially mobilizeable connections in the network, reflecting the depth of the system's exploration of embedded trust resources. The average path shortening rate is also applied as an auxiliary metric to measure the compression ratio of the shortest propagation path of key health information from the initiating node to the target group in the social network, thereby assessing the effect of resource guidance on improving overall network fluidity. Experiments are conducted under two institutional environments: weak incentives (low resource supply and weak individual response

tendency) and strong incentives (high resource redundancy and high willingness to collaborate), comparing the performance of the methods. Thirty independent simulations are performed for each configuration. The relationship call efficiency and topology optimization capabilities of each method in a dynamic interactive structure are statistically analyzed, revealing the essential differences in the social capital conversion efficiency of different decision-making mechanisms. Table 2 shows the effectiveness of different resource allocation strategies in activating social capital and improving social network mobility under weak and strong incentives. The results show that the proposed MADRL model significantly outperforms the comparative methods in both scenarios. Under the weak incentive system, the social capital activation rate is $79.3\% \pm 3.6\%$, and the average path shortening rate is $61.8\% \pm 5.0\%$. This not only achieves a higher relationship chain activation rate but also performs exceptionally well in terms of average path shortening rate, demonstrating its ability to effectively identify and leverage key social connections, promoting the more efficient diffusion of health intervention information and supportive behaviors within the network. This advantage stems from the collaborative design of a dynamic social embedding mechanism and hierarchical credit allocation, which enables the agent to precisely locate high-influence nodes and guide the actual activation of potential relationships through incentive-compatible strategies. In contrast, traditional MADRL, due to its lack of heterogeneous role modeling, struggles to distinguish between core supporters and marginal individuals. Linear programming and greedy algorithms, limited by static topological assumptions, cannot capture the temporal evolution of interaction potential. Random allocation exhibits the least structural sensitivity. Notably, under weak incentives, the performance of various methods generally declines, while the standard deviation of the proposed model remains consistently low.

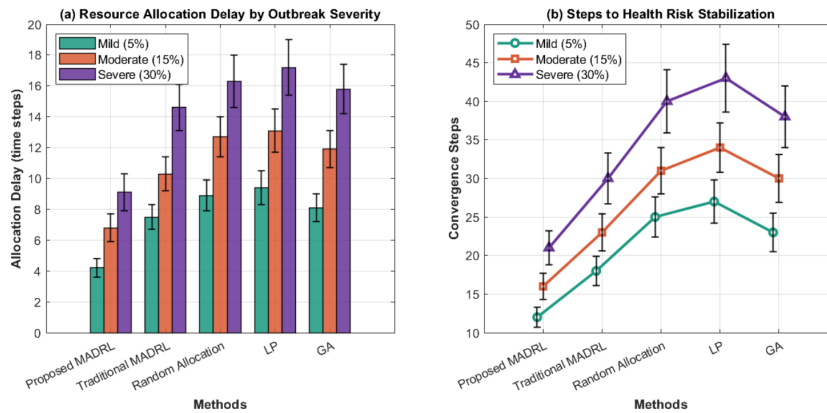


FIGURE 7. Response efficiency and system stability

TABLE 2. Social capital activation rate and average path shortening rate

Method	Condition	Activation Rate (%)	Mean Path Shortening Ratio (%)
Proposed MADRL	Strong	86.4 ± 3.1	68.2 ± 4.5
Traditional MADRL	Strong	72.1 ± 4.7	51.3 ± 5.9
Random Allocation	Strong	43.7 ± 6.8	28.6 ± 7.2
LP	Strong	50.2 ± 5.3	33.8 ± 6.1
GA	Strong	58.9 ± 5.9	40.5 ± 6.7
Proposed MADRL	Weak	79.3 ± 3.6	61.8 ± 5.0
Traditional MADRL	Weak	63.5 ± 5.2	44.7 ± 6.4
Random Allocation	Weak	38.2 ± 7.1	25.1 ± 7.8
LP	Weak	44.6 ± 5.8	30.2 ± 6.9
GA	Weak	51.4 ± 6.3	36.9 ± 7.3

IV. CONCLUSIONS

This paper addresses the core challenges of social capital quantification and inefficient multi-agent collaboration in community health management by constructing an optimal resource allocation model that integrates dynamic social network embedding and hierarchical multi-agent deep reinforcement learning. This methodology offers a powerful parallel for the textile industry, where it could be adapted to address the core challenges of quantifying supplier relational strength and optimizing the often-inefficient collaboration among multiple manufacturing agents by constructing an optimal production and material allocation model. A temporal graph neural network is used to continuously represent resident interactions, transforming abstract social capital into computable node embedding features. A heterogeneous layered intelligent agent architecture is designed to give individuals and organizations differentiated observation and action spaces, truly recreating the role division in real governance. A joint reward mechanism encompassing health gains, resource efficiency, and network fairness is applied, and local credit allocation is performed in conjunction with counterfactual baselines to improve the accuracy and interpretability of policy learning. Policy optimization is

achieved within a centralized training and distributed execution framework using the MAPPO algorithm, enhancing the system's convergence stability in high-dimensional dynamic environments. Experiments demonstrate that the model significantly outperforms traditional methods in terms of demand-normalized allocation equity, health gains, social capital utilization, and multi-agent collaborative consistency, demonstrating particularly good adaptability and robustness in communities of varying densities and under epidemic impact scenarios. This research not only constructs a complete closed loop from perceiving social structures to generating collaborative decisions, but also provides a referable solution for intelligent resource allocation in complex social governance. This achievement holds significant promise beyond social systems; the capability for end-to-end optimization, linking supplier network structure perception to synchronized output decisions, offers a direct reference for intelligent resource allocation and collaborative planning within complex industrial supply chain scenarios, such as the textile industry.

AUTHOR CONTRIBUTIONS

Conceptualization – Li Jin; methodology – Li Jin; investigation – Li Jin; writing-original draft preparation – Li Jin. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics Committee of Shanghai Xinnasongshan Biopharmaceutical Technology Co., Ltd, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] M. Zhang, M. Zhao, and X. Wei, "Research progress on community health management model for older adults with chronic diseases and multiple morbidities," *British Journal of Hospital Medicine*, vol. 85, no. 10, pp. 1-9, 2024, doi: 10.12968/hmed.2024.0256.
- [2] Y. Jiang and F. Zhen, "The role of community service satisfaction in the influence of community social capital on the sense of community belonging: a case study of Nanjing, China," *Journal of Housing and the Built Environment*, vol. 37, no. 2, pp. 705-721, 2022, doi: 10.1007/s10901-021-09862-9.
- [3] D. M. Littman, "Third places, social capital, and sense of community as mechanisms of adaptive responding for young people who experience social marginalization," *American Journal of Community Psychology*, vol. 69, no. 3-4, pp. 436-450, 2022, doi: 10.1002/ajcp.12531.
- [4] M. S. U. Rahman, D. Simmons, M. C. Shone, and N. N. Ratna, "Social and cultural capitals in tourism resource governance: the essential lenses for community focussed co-management," *Journal of Sustainable Tourism*, vol. 30, no. 11, pp. 2665-2685, 2022, doi: 10.1080/09669582.2021.1903016.
- [5] D. Wang and S. Li, "Social conflicts and their resolution paths in the commercialized renewal of old urban communities in China under the perspective of public value," *Journal of Urban Management*, vol. 14, no. 2, pp. 402-417, 2025, doi: 10.1016/j.jum.2024.11.007.
- [6] S. Amri and J. Sihotang, "Impact of Poverty Reduction Programs on Healthcare Access in Remote Ar-eas: Fostering Community Development for Sustainable Health," *Law and Economics*, vol. 17, no. 3, pp. 170-185, 2023, doi: 10.35335/laweco.v17i3.43.
- [7] B. Shin, "Exploring network measures of social capital: Toward more relational measurement," *Journal of Planning Literature*, vol. 36, no. 3, pp. 328-344, 2021, doi: 10.1177/0885412221999415.
- [8] S. Li, X. He, and J. Chen, "Exploring the relationship between interaction patterns and social capital accumulation in connectivist learning," *Interactive Learning Environments*, vol. 32, no. 6, pp. 2691-2712, 2024, doi: 10.1080/10494820.2022.2157839.
- [9] M. M. Li, K. Huang, and M. Zitnik, "Graph representation learning in biomedicine and healthcare," *Nature biomedical engineering*, vol. 6, no. 12, pp. 1353-1369, 2022, doi: 10.1038/s41551-022-00942-x.
- [10] S. Anwar, M. O. Beg, K. Saleem, Z. Ahmed, A. R. Javed, and U. Tariq, "Social relationship analysis using state-of-the-art embeddings," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 22, no. 5, pp. 1-21, 2023, doi: 10.1145/3539608.
- [11] H. Xu, C. Zhang, and Y. Huang, "Social trust, social capital, and subjective well-being of rural residents: Microempirical evidence based on the Chinese General Social Survey (CGSS)," *Humanities and Social Sciences Communications*, vol. 10, no. 1, pp. 1-13, 2023, doi: 10.1057/s41599-023-01532-1.
- [12] H. Hwang and Y. Lee, "Community social capital, racial diversity, and philanthropic resource mobilization in the time of a pandemic," *City & Community*, vol. 22, no. 1, pp. 22-47, 2023, doi: 10.1177/15356841221119181.
- [13] N. S. Sabet and S. Khaksar, "The performance of local government, social capital and participation of villagers in sustainable rural development," *The Social Science Journal*, vol. 61, no. 1, pp. 1-29, 2024, doi: 10.1080/03623319.2020.1782649.
- [14] T. Baycan and Ö. Öner, "The dark side of social capital: a contextual perspective," *The Annals of Regional Science*, vol. 70, no. 3, pp. 779-798, 2023, doi: 10.1007/s00168-022-01112-2.
- [15] L. K. Ozanne and J. L. Ozanne, "Disaster recovery: How ad hoc marketing systems build and mobilize social capital for service delivery," *Journal of Public Policy & Marketing*, vol. 40, no. 3, pp. 372-388, 2021, doi: 10.1177/07439156211000355.
- [16] I. G. Okpechi, A. K. Bello, V. A. Luyckx, N. Wearne, C. R. Swanepoel, and V. Jha, "Building optimal and sustainable kidney care in low resource settings: the role of healthcare systems," *Nephrology*, vol. 26, no. 12, pp. 948-960, 2021, doi: 10.1111/nep.13935.
- [17] L. K. Mayger and C. D. Hochbein, "Growing connected: Relational trust and social capital in community schools," *Journal of Education for Students Placed at Risk (JESPAR)*, vol. 26, no. 3, pp. 210-235, 2021, doi: 10.1080/10824669.2020.1824676.
- [18] M. Romakkaniemi, T. Harrikari, S. Saraniemi, L. Tiitinen, and V. Fiorentino, "'Bonding, bridging and linking the last resort tailboard': Shifts in social workers' professional positions and mobilizing adaptive capital during the coronavirus pandemic," *Nordic Social Work Research*, vol. 13, no. 3, pp. 366-379, 2023, doi: 10.1080/2156857X.2021.1992489.
- [19] S. Kavanagh, A. Shiell, P. Hawe, and K. Garvey, "Resources, relationships, and systems thinking should inform the way community health promotion is funded," *Critical Public Health*, vol. 32, no. 3, pp. 273-282, 2022, doi: 10.1080/09581596.2020.1813255.
- [20] A. Olaniran, B. Madaj, S. Bar-Zeev, A. Banke-Thomas, and N. van den Broek, "Factors influencing motivation and job satisfaction of community health workers in Africa and Asia—A multi-country study," *The International journal of health planning and management*, vol. 37, no. 1, pp. 112-132, 2022, doi: 10.1002/hpm.3319.