Learning to Diffuse: Costly Communication in Networked Mechanism Design

Sebastiano Messina SEBASTIANO.MESSINA@TUWIEN.AC.AT
Tatjana Chavdarova TATJANA.CHAVDAROVA@TUWIEN.AC.AT
TU Wien, Austria

Abstract

Multi-agent systems often need to route tasks through networks of agents, but communication is itself costly and strategic. The so-called *diffusion auctions* formalize endogenous participation in such networks, yet most models treat forwarding as costless and study one-shot mechanisms. We introduce the *Learning Diffusion Mechanism* (LDM), a repeated Network Vickrey–Clarke–Groves (VCG) framework in which informed agents observe private task values, choose which neighbors to inform, and receive their VCG utility minus a per-link propagation cost. Additive costs preserve truthful bidding, reducing the strategic problem to learning valuation-dependent diffusion policies. We characterize the induced Bayesian diffusion game, establish existence of a stationary Bayes-Nash equilibrium, and identify a welfare-aligned algebraic potential structure for the associated mean-field learning dynamics. We also derive a model-free policy-gradient update that uses only realized utilities. This work connects *diffusion auctions* (from mechanism design) and agentic systems, showing how local communication costs can shrink the realized auction network even when allocation is efficient within the reached subgraph.

1. Introduction

Networks of agents, including large language models (LLMs), are increasingly used to solve complex tasks. A coordinator typically receives a task and decides which agent should handle it. In practice, however, the coordinator may know or be able to contact only a small fraction of the available agents. The best agent for a task may be reachable only through intermediate agents that can forward, refer, or delegate the request. Such communication has costs: it may consume tokens, time, bandwidth, attention, privacy budget, or API resources. In addition, an intermediary may have little incentive to reveal a stronger downstream agent, since doing so can reduce its own chance of being selected. The challenge is therefore not only to identify the best agent, but also to design coordination mechanisms that encourage useful delegation while keeping communication costs low.

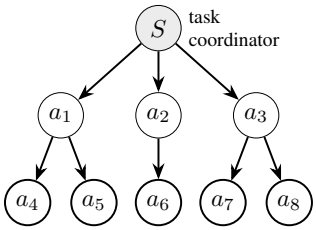


Figure 1: A diffusion network for task routing. Coordinator S initially reaches a subset of agents; downstream agents participate if upstream agents forward the task.

Diffusion auctions are a natural framework for this setting because they model markets in which information sharing is itself strategic (Emek et al., 2011; Jia et al., 2023; Li et al., 2017; Zhang et al., 2024). A *seller*—here, the task coordinator—directly reaches only a small set of *buyers*—or agents—in a network. These agents may forward the task to their neighbors, who may then inform further agents, creating a tree-like communication structure; see Figure 1. Participation is

© S. Messina & T. Chavdarova.

consequently endogenous: an agent can compete for the task only if upstream agents activate the communication links that reach it. Because informing additional agents may introduce stronger competitors, strategic agents may withhold information unless the mechanism rewards their contribution to expanding the reachable network.

Classical diffusion-auction mechanisms study this tension in social and economic networks, including referrals, recruitment, online marketplaces, and advertising (e.g., Emek et al., 2011; Li et al., 2017). Their objective is twofold: to elicit agents' private values truthfully and to incentivize them to propagate the auction. Subsequent work has studied revenue, privacy, and structural extensions of diffusion auctions (Huang et al., 2025; Jia et al., 2023; Zhang et al., 2024). We build on the Network *Vickrey–Clarke–Groves* (VCG) mechanism (Clarke, 1971; Groves, 1973; Vickrey, 1961). Given the realized set of reachable agents, VCG allocates the item to the highest-valued agent, which directly matches our task-routing objective of assigning a request to the best available executor. Its marginal-contribution payments can also reward agents that are essential for reaching the selected executor. These properties make VCG a natural starting point for studying strategic task propagation in agentic networks. Existing diffusion-auction models commonly assume that propagation is costless. Under this assumption, expanding participation can be made individually rational through suitable diffusion rewards. Our focus is the regime suggested above: activating an information link has a direct local cost. Full propagation is then no longer automatically optimal. Instead, each agent must decide which neighbors are worth informing, balancing the potential benefit of reaching a high-value downstream agent against the cost of expanding the communication tree.

We model this setting as a repeated strategic diffusion problem[1]. In each round, agents observe private task values, submit bids, choose which neighbors to inform, and receive their VCG payoff minus a marginal cost for each activated outgoing link. Since the propagation cost is additive and independent of the bid, truthful bidding remains weakly dominant for any fixed diffusion decision. The strategic problem then reduces to learning type-dependent diffusion policies: low-value agents may act as information brokers, while high-value agents may strategically truncate propagation to avoid inviting stronger competitors. This turns diffusion auctions into a *learning problem over strategic networks*. Unlike standard learning-based auction design, which typically assumes a fixed set of bidders (Duetting et al., 2019; Ivanov et al., 2022; Rahme et al., 2021), the participant set, communication pattern, and allocation outcome here are jointly endogenous. The VCG structure is useful for this learning problem because payoffs are marginal contributions to welfare. Under the additive cost model, this yields an algebraic potential structure that aligns unilateral utility changes with changes in the penalized objective used below.

A closely related line of work studies one-shot diffusion mechanisms in intermediated economic networks with transaction costs incurred when a selected trade traverses network edges. These mechanisms design transfers to recover truthful sharing, efficient allocation, or budget feasibility despite such costs (Li et al., 2018, 2024a; Luo and Wang, 2026); see Appendix A. In contrast, we model the cost of *information propagation itself*: agents incur a cost when they activate communication links, before the final allocation and trading path are known. Rather than designing a one-shot mechanism that restores full dissemination, we study a repeated environment in which agents learn

1. Herein, "diffusion" refers to strategic information propagation; not to diffusion models for generative modeling.

selective, valuation-dependent diffusion policies from realized utilities. To the best of our knowledge, prior work has not studied repeated diffusion auctions in which agents strategically incur per-link information-propagation costs and adapt their forwarding decisions over time from payoff feedback. Beyond auctions, the same abstraction captures a basic challenge in emerging multi-agent systems: agents must learn when to communicate, reveal information, update shared state, or delegate when communication entails computational, privacy, or strategic costs. Examples include multi-agent software engineering, collaborative memory construction, autonomous-mobility coordination, and decentralized electricity-market agents (e.g., Foerster et al., 2018; Hong et al., 2024; Wu et al., 2023; Yao et al., 2023).

Contributions. We introduce the *Learning Diffusion Mechanism* (LDM), a repeated Network-VCG framework for strategic diffusion. Our main contributions are:

- We formalize repeated allocation with per-link diffusion costs and show that truthful bidding remains weakly dominant, reducing strategic behavior to diffusion decisions;
- We characterize the resulting Bayesian diffusion game, show stationary Bayes-Nash equilibrium existence, and identify an algebraic potential structure for the mean-field learning dynamics; and
- We derive a model-free policy-gradient update through which agents adapt valuation-dependent diffusion propensities using only realized net utilities, without complete knowledge of the network topology or other agents' valuation distributions.

An illustrative example and learning simulation are deferred to Appendices E and F, respectively. We additionally discuss an exploratory extension to the Information Diffusion Mechanism (IDM) in Appendix G. IDM is a common diffusion-auction benchmark, but its sequential threshold allocation rule can select an upstream intermediary rather than the highest-valued reachable agent and does not directly inherit the VCG-based potential structure developed in this work.

2. Setting & Preliminaries

Let $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ be a directed communication network over buyer-agents $\mathcal{N} = \{1, \ldots, n\}$. A seller, or task coordinator, s holds one item, interpreted as a task, with reserve price zero. The seller initially informs a fixed subset of agents; we treat these invitations as fixed active links. For each agent i, let $r_i = \{j \in \mathcal{N} : (i, j) \in \mathcal{E}\}$ be the neighbors that i can inform. Each agent i has a private value $v_i \geq 0$. An agent can participate only after receiving the auction information. Once informed, agent i submits $a_i = (b_i, \mathbf{x}_i)$, where $b_i \in \mathbb{R}_+$ is its bid. The vector $\mathbf{x}_i = (x_{i,j})_{j \in r_i} \in \{0, 1\}^{|r_i|}$ is its diffusion action, with $x_{i,j} = 1$ if i informs neighbor j. The joint diffusion profile $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ induces an active graph $\mathcal{G}_{\mathbf{x}}$. Let $\mathcal{R}(\mathbf{x})$ be the agents reachable from s in this graph. Only agents in $\mathcal{R}(\mathbf{x})$ submit bids and compete for the item.

Our baseline is the costless one-shot Network VCG mechanism. Given $(\mathbf{b}, \mathbf{x})$, VCG allocates the item to a highest-bidding reachable agent, $w \in \arg\max_{j \in \mathcal{R}(\mathbf{x})} b_j$, with deterministic tie-breaking. It also assigns each reachable agent a gross payoff $u_i^{\text{vcg}}(\mathbf{b}, \mathbf{x})$. This payoff is the agent's VCG marginal contribution. The winner obtains allocation utility net of its payment. A critical intermediary may be rewarded when the winner is reachable only through that intermediary. Appendix B gives the formal definitions of reachability, critical diffusion nodes, payments, and the piecewise expression for u_i^{vcg}.

3. Learning Diffusion Mechanism (LDM): Repeated Game

We next extend the costless one-shot VCG baseline to a repeated selling environment with information propagation costs. Time is discrete, $t = 1, \ldots, T$. In each round, agents are given a new valuation draw and choose both a bid and a diffusion action. Unlike the costless VCG, we introduce an exogenous marginal propagation cost $c > 0$; each agent i pays cost c for every neighbor to whom it sends the auction information. Proofs are deferred to Appendix D.

At round t, agent i observes a private valuation $v_i(t) \geq 0$, independently drawn from an absolutely continuous distribution F_i, with bounded support $[0, v_{max}]$. After informed agents simultaneously submit actions $a_i(t) = (b_i(t), \mathbf{x}_i(t))$, the mechanism runs VCG on the realized diffusion graph $\mathcal{G}_{\mathbf{x}}$ and bids $\mathbf{b}(t)$. The realized net utility $u_i(t)$ of agent i is additively separable:

$$u_i(t) = u_i^{vcg}(\mathbf{b}(t), \mathbf{x}(t)) - c \sum_{j \in r_i} x_{i,j}(t) \,. \tag{u_i}$$

3.1. Strategy Space Reduction and Expected Utility

The marginal cost $c > 0$ breaks the dominant strategy of full diffusion. However, the additive separability of Equation (u_i) allows for an analytical separation between bidding and diffusion strategies.

Proposition 1 (truthful bidding persistence). *Under VCG with propagation costs $c > 0$, truthful bidding, $(b_i(t) = v_i(t))$, is a weakly dominant strategy for any diffusion profile $\mathbf{x}_i(t)$.*

By proposition 1, equilibrium analysis can be restricted to the diffusion phase. Let $\mathbf{X}_i = \{0,1\}^{|r_i|}$ be the finite set of pure diffusion actions available to agent i. A behavioral strategy is defined as a measurable function $\sigma_i : [0, v_{max}] \to \Sigma_i = \Delta(\mathbf{x}_i)$, where $\sigma_i(\mathbf{x}_i|v_i)$ denotes the probability of selecting the pure action $\mathbf{x}_i \in \mathbf{X}_i = \{0,1\}^{|r_i|}$ conditional on their realized valuation v_i.

Given the joint strategic profile of the remaining agents σ_{-i}, the interim expected utility in Equation (U_i) for an agent of type v_i adopting a pure action $\mathbf{x}_i$ is obtained by averaging ex-post utility over other agents' valuations and diffusion actions:

$$\begin{aligned} U_i(\mathbf{x}_i, \sigma_{-i} \mid v_i) &= \mathbb{E}_{\mathbf{v}_{-i} \sim F_{-i}} \left[\mathbb{E}_{\mathbf{x}_{-i} \sim \sigma_{-i}(\cdot|\mathbf{v}_{-i})} \left[u_i((v_i, \mathbf{v}_{-i}), (\mathbf{x}_i, \mathbf{x}_{-i})) \right] \right] \\ &= \int_{\mathbf{v}_{-i}} \sum_{\mathbf{x}_{-i} \in \mathbf{X}_{-i}} u_i((v_i, \mathbf{v}_{-i}), (\mathbf{x}_i, \mathbf{x}_{-i})) \left(\prod_{j \neq i} \sigma_j(\mathbf{x}_j|v_j) \right) dF_{-i}(v_{-i}) \,. \end{aligned} \tag{U_i}$$

To characterize equilibrium, we use the best-response mapping in Equation ($\mathcal{B}_i$).

Definition 2 (best-response correspondence). *The best-response mapping $\mathcal{B}_i \colon \Sigma_{-i} \to \Sigma_i$ assigns to each σ_{-i} the behavioral strategies that maximize agent i's expected utility for almost every type:*

$$\mathcal{B}_i(\sigma_{-i}) = \left\{ \sigma_i \in \Sigma_i \mid supp(\sigma_i(\cdot|v_i)) \subseteq \arg\max_{\mathbf{x}_i \in \mathbf{X}_i} U_i(\mathbf{x}_i, \sigma_{-i} \mid v_i) \textit{ for a.e. } v_i \right\} . \tag{$\mathcal{B}_i$}$$

To analyze strategic diffusion decisions, we define the penalized social welfare in Equation (Φ). It is evaluated at true valuations and subtracts all propagation costs incurred by the network:

$$\Phi(\mathbf{v}, \mathbf{x}) = W(\mathbf{v}, \mathbf{x}) - c \sum_{k \in \mathcal{N}} \sum_{l \in r_k} x_{k,l} = \max_{j \in \mathcal{R}(\mathbf{x})} v_j - c \sum_{k \in \mathcal{N}} \sum_{l \in r_k} x_{k,l} \,. \tag{Φ}$$

Lemma 3 (utility identity). *The ex-post realized net utility of any agent $i \in \mathcal{N}$ satisfies the following identity:*

$$u_i(\mathbf{v}, \mathbf{x}) = \Phi(\mathbf{v}, \mathbf{x}) - \Phi_{-i}(\mathbf{v}, \mathbf{x}_{-i}) \,. \tag{u_i^Φ}$$

Theorem 4 (algebraic potential identity). *Under the LDM payoff definition, unilateral deviations satisfy*

$$u_i(v, x_i', x_{-i}) - u_i(v, x_i, x_{-i}) = \Phi(v, x_i', x_{-i}) - \Phi(v, x_i, x_{-i}).$$

Thus, Φ, Equation (Φ), acts as an exact algebraic potential for the normal-form diffusion game.

Remark 5 (interpretation of the counterfactual potential). *The counterfactual term Φ_{-i} in Equation (u_i^Φ) is an algebraic baseline, not a literal welfare value after removing agent i. Its welfare term removes the downstream branch $d_i(\mathbf{x})$, matching the VCG externality, but its cost term keeps the diffusion costs of all agents $k \neq i$ under the original profile. It may therefore count costs for agents who would not be reached in the truncated network. This is the main limitation of the potential construction: it gives exact alignment for unilateral deviations, but Φ_{-i} should not be interpreted as feasible social welfare. Its role is only to keep the baseline independent of $\mathbf{x}_i$, so that $\Phi - \Phi_{-i}$ isolates agent i's net utility.*

Having formally reduced the game to the diffusion decisions conditional on types (Proposition 1) and established the exact potential structure of the model, we now establish the existence of a stationary equilibrium for the network.

Theorem 6 (existence of stationary Bayes-Nash equilibrium). *For finite network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ and a marginal cost $c > 0$, if valuation distributions $\{F_i\}_{i=1}^n$ are absolutely continuous within support $[0, v_{\max}]$, then the LDM admits at least one stationary Bayes-Nash Equilibrium (BNE).*

4. Model-Free Learning of Diffusion Policies

In networked environments characterized by incomplete information, agents lack prior knowledge of the underlying network topology and the valuation distributions of competing participants. Consequently, calculating the exact expected utility of diffusing information to a neighbor, and thereby deriving an analytical best-response mapping, becomes computationally intractable. To resolve this, agents must rely on model-free learning algorithms, leveraging their observed realized net utility $u_i(t)$ to iteratively adjust and optimize their behavioral strategies. This framework formally maps the static diffusion auction into a Multi-Agent Reinforcement Learning (MARL) paradigm governed by independent policy gradient dynamics.

4.1. Parametrized Diffusion Policies and Gradient Updates

To establish a rigorous mathematical foundation, we parameterize the behavioral diffusion strategy of agent i by a vector $\theta_i \in \mathbb{R}^{|r_i|}$. Because the optimal diffusion strategy depends on the agent's competitive standing, this parameterization must be conditioned on the agent's valuation type $v_i(t)$.

Remark 7. *Under VCG mechanism, an agent's incentive to propagate information is the potential reward extracted by acting as a critical diffusion node for a higher-valuation winner. Conversely, for an agent with a high valuation $v_i(t)$, further diffusion decreases expected utility by introducing competitors that could either outbid the agent or increase their required VCG payment. Consequently, the optimal diffusion strategy is strictly type-dependent; an agent's propensity to diffuse, $\theta_{i,j}$ must be conditioned on $v_i(t)$. This structural tension induces bifurcated strategic behavior: low-valuation agents rationally act as information brokers to capture diffusion rewards, whereas high-valuation agents strategically truncate network propagation to suppress local competition.*

The propensity of agent i to activate an outgoing edge to neighbor $j \in r_i$ is governed by the parameter $\theta_{i,j}(v_i)$. We model the independent activation probability of each edge using the sigmoid map in Equation (P_{ij}):

$$\mathbb{P}_{i,j}(v_i;\theta_{i,j}) = \mathbb{P}_{i,j}(\theta_{i,j}(v_i)) = \frac{1}{1+e^{-\theta_{i,j}(v_i)}}\,. \tag{P_{ij}}$$

Assuming conditional independence across localized routing decisions, the joint policy in Equation (σ_i) is a product of Bernoulli distributions:

$$\sigma_i(\mathbf{x}_i|v_i;\theta_i) = \prod_{j\in r_i} \left[\mathbb{P}_{i,j}(\theta_{i,j}(v_i))\right]^{x_{i,j}} \left[1-\mathbb{P}_{i,j}(\theta_{i,j}(v_i))\right]^{1-x_{i,j}}\,. \tag{σ_i}$$

Let $\theta = (\theta_1,\ldots,\theta_n)$ denote the joint strategy parameter profile. The ex-ante expected objective in Equation (J_i) integrates ex-post utility over valuations and induced stochastic policies:

$$J_i(\theta) = \mathbb{E}_{\mathbf{v}\sim F}\left[\sum_{\mathbf{x}\in\mathbf{X}} u_i(\mathbf{v},\mathbf{x})\left(\prod_{k\in\mathcal{N}} \sigma_k(\mathbf{x}_k|v_k;\theta_k)\right)\right]. \tag{J_i}$$

To maximize J_i, agents use REINFORCE (Sutton et al., 1999; Williams, 1992). For edge (i,j), the one-sample gradient estimator is $\hat{g}_{i,j}(t) = u_i(t)(x_{i,j}(t) - \mathbb{P}_{i,j}(t))$. This yields the stochastic gradient ascent update in Equation (PG):

$$\theta_{i,j}(t+1) = \theta_{i,j}(t) + \alpha_t u_i(t)(x_{i,j}(t) - \mathbb{P}_{i,j}(t)), \tag{PG}$$

where $\alpha_t \in (0,1)$ is the learning rate at time step t. Algorithm 1 summarizes the resulting decentralized LDM update; the stochastic-gradient derivation is given in Appendix C.

4.2. Stochastic Approximation and the Mean-Field Limit

To rigorously analyze the convergence properties of this decentralized MARL system, we map the discrete stochastic updates to a deterministic continuous-time trajectory using Stochastic Approximation theory. We decompose the discrete update in Equation (PG) into the exact analytical gradient and a zero-mean noise term:

$$\theta_{i,j}(t+1) = \theta_{i,j}(t) + \alpha_t\left[\frac{\partial J_i(\theta(t))}{\partial\theta_{i,j}} + M_{i,j}(t+1)\right]. \tag{SA}$$

Proposition 8 (martingale difference noise). *Let $\mathcal{F}_t = \sigma(\theta(0), \{\mathbf{v}(\tau),\mathbf{x}(\tau)\}_{\tau\le t})$ denote the natural filtration representing the historical σ-algebra of the process up to time t. The error sequence $M_{i,j}(t+1) = \hat{g}_{i,j}(t) - \frac{\partial J_i(\theta(t))}{\partial\theta_{i,j}}$ is a martingale difference sequence adapted to $\mathcal{F}_t$.*

Theorem 9 (mean-field tracking). *Assume the learning rate sequence $\{\alpha_t\}$ satisfies the Robbins-Monro conditions for infinite travel and noise suppression as follows:*

$$\sum_{t=1}^{\infty} \alpha_t = \infty\,, \qquad \text{(A1)} \qquad\qquad \sum_{t=1}^{\infty} \alpha_t^2 < \infty\,. \qquad \text{(A2)}$$

As $t \to \infty$, the discrete-time stochastic process $\theta(t)$ asymptotically tracks the solution trajectories of the continuous-time mean-field ODE in Equation (ODE):

$$\frac{d\theta_i}{dt} = \nabla_{\theta_i} J_i(\theta)\,. \qquad \text{(ODE)}$$

Under the step-size conditions, the cumulative sum of the martingale difference noise asymptotically vanishes almost surely. Therefore, the decentralized learning dynamics are asymptotically described by a smooth vector field defined by the expected analytical gradients.

4.3. Potential Game Dynamics and Asymptotic Stability

With the continuous-time dynamics established, we evaluate the system's global behavior. The expected global social welfare in Equation ($\Phi(\theta)$) evaluates the potential function $\Phi(\mathbf{v}, \mathbf{x})$ over the joint type distribution and current parameterized policies:

$$\Phi(\theta) = \mathbb{E}_{\mathbf{v}\sim F}\left[\sum_{\mathbf{x}\in\mathbf{X}} \Phi(\mathbf{v}, \mathbf{x}) \left(\prod_{k\in\mathcal{N}} \sigma_k(\mathbf{x}_k | v_k; \theta_k)\right)\right]. \qquad (\Phi(\theta))$$

Similarly, let $\Phi_{-i}(\theta_{-i})$ denote the expected counterfactual potential function under exclusion of i.

Theorem 10 (asymptotic convergence of LDM dynamics). *Let $\theta(t)$ denote the joint parameter vector governing the network diffusion policies. Under continuous-time gradient ascent dynamics, the system's trajectory monotonically increases the expected global social welfare $\Phi(\theta)$ in Equation ($\Phi(\theta)$) and converges asymptotically to a stationary point.*

5. Discussion

LDM makes a simple point precise: in networked allocation problems, once agents pay to activate links, the auction no longer runs on the underlying full network, but on the network that agents strategically choose to reveal. Additive propagation costs preserve truthful bidding, so strategic behavior concentrates in the diffusion policy. This turns diffusion auctions into a learning problem over endogenous participation, unlike standard learning-based auction design with fixed bidder sets. Under Network VCG, the marginal-contribution payoff structure gives a potential-game view of the resulting learning dynamics, while the counterfactual potential should be read as an algebraic device rather than a literal feasible welfare counterfactual. The broader message is that costly communication can make efficient local allocation coexist with sparse global participation. Agents may learn not to forward information, not because downstream agents are irrelevant, but because inviting them is privately costly or competitively dangerous. This lens also applies beyond auctions: routing, delegation, memory sharing, and disclosure in agentic systems are strategic diffusion decisions when communication has cost (Foerster et al., 2018; Hong et al., 2024; Wu et al., 2023; Yao et al., 2023). Next steps are to condition diffusion on task features, study revenue and welfare under learned selective diffusion, and extend the framework to multi-unit or combinatorial diffusion auctions (Fang et al., 2023; Li et al., 2024b).

Acknowledgments

Authors are thankful for the generous support of the Vienna Science and Technology Fund (WWTF), grant ID: 10.47379/VRG25019.

References

Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971. doi: 10.1007/BF01726210.

Michael J. Curry, Uro Lyi, Tom Goldstein, and John P. Dickerson. Learning revenue-maximizing auctions with differentiable matching. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 6062–6073. PMLR, 28–30 Mar 2022.

Paul Duetting, Zhe Feng, Harikrishna Narasimhan, David Parkes, and Sai Srivatsa Ravindranath. Optimal auctions through deep learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1706–1715. PMLR, 09–15 Jun 2019.

Yuval Emek, Ron Karidi, Moshe Tennenholtz, and Aviv Zohar. Mechanisms for multi-level marketing. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, EC '11, page 209–218. Association for Computing Machinery, 2011. ISBN 9781450302616. doi: 10.1145/1993574.1993606.

Yuan Fang, Mengxiao Zhang, Jiamou Liu, Bakh Khoussainov, and Mingyu Xiao. Multi-unit auction over a social network. *arXiv:2302.08924*, 2023.

Jakob Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 122–130, 2018.

Irving Leonard Glicksberg. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society*, 3 (1):170–174, 1952. doi: 10.2307/2032478.

Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973. doi: 10.2307/1914085.

Wenshuo Guo, Michael Jordan, and Emmanouil Zampetakis. Robust learning of optimal auctions. In *Advances in Neural Information Processing Systems*, 2021.

Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. MetaGPT: Meta programming for a multi-agent collaborative framework. In *The Twelfth International Conference on Learning Representations*, 2024.

Kexin Huang, Ziqian Chen, Xue Wang, Chongming Gao, Jinyang Gao, Bolin Ding, and Xiang Wang. Auctionformer: A unified deep learning algorithm for solving equilibrium strategies in auction games. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 19635–19659. PMLR, 21–27 Jul 2024.

Yifan Huang, Dong Hao, Zhiyi Fan, Yuhang Guo, and Bin Li. Approximate revenue maximization for diffusion auctions. *arXiv:2507.14470*, 2025.

Dmitry Ivanov, Iskander Safiulin, Igor Filippov, and Ksenia Balabaeva. Optimal-er auctions through attention. In *Advances in Neural Information Processing Systems*, 2022.

Fengjuan Jia, Mengxiao Zhang, Jiamou Liu, and Bakh Khoussainov. Incentivising diffusion while preserving differential privacy. In *The 39th Conference on Uncertainty in Artificial Intelligence*, 2023.

Bin Li, Dong Hao, Dengji Zhao, and Tao Zhou. Mechanism design in social networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Bin Li, Dong Hao, Dengji Zhao, and Tao Zhou. Customer sharing in economic networks with costs. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 368–374, 2018.

Bin Li, Dong Hao, and Dengji Zhao. Diffusion auction design with transaction costs. *Autonomous Agents and Multi-Agent Systems*, 38(1), 2024a. ISSN 1387-2532. doi: 10.1007/s10458-023-09631-8.

Xuanyu Li, Miao Li, Yuhan Cao, and Dengji Zhao. Combinatorial diffusion auction design. *arXiv:2410.22765*, 2024b.

Yunpeng Luo and Yisong Wang. Design of decentralized diffusion auction mechanism based on intermediated market. *Journal of Computer Applications*, 2026. doi: 10.11772/j.issn.1001-9081.2026010052.

Paul R Milgrom and Robert J Weber. Distributional strategies for games with incomplete information. *Mathematics of operations research*, 10(4):619–632, 1985.

Neehar Peri, Michael Curry, Samuel Dooley, and John Dickerson. Preferencenet: Encoding human preferences in auction design with deep learning. In *Advances in Neural Information Processing Systems*, volume 34, pages 17532–17542. Curran Associates, Inc., 2021.

Jad Rahme, Samy Jelassi, and S. Matthew Weinberg. Auction learning as a two-player game. In *International Conference on Learning Representations*, 2021.

Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999.

Vasilis Syrgkanis. A sample complexity measure with applications to learning optimal auctions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961. doi: 10.1111/j.1540-6261.1961.tb02789.x.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992. ISSN 0885-6125. doi: 10.1007/BF00992696.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation. *arXiv:2308.08155*, 2023.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023.

Tianyi Zhang, Dengji Zhao, Junyu Zhang, and Sizhe Gu. Cost sharing under private costs and connection control on directed acyclic graphs. DAI '23, 2023. ISBN 9798400708480. doi: 10.1145/3627676.3627689.

Yao Zhang, Shanshan Zheng, and Dengji Zhao. Optimal diffusion auctions. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '24, page 2600–2602. International Foundation for Autonomous Agents and Multiagent Systems, 2024. ISBN 9798400704864.

Appendix A. Related Work

Diffusion and network-aware mechanism design. Classical auction theory studies strategic allocation under incomplete information, typically taking the set of participating agents as fixed (Milgrom and Weber, 1985). A complementary line of work considers mechanisms in which agents can affect participation itself by forwarding information through a social network. Early work on multi-level marketing formalized reward mechanisms for incentivizing propagation in networks (Emek et al., 2011). Diffusion auctions bring this idea into auction design: the seminal mechanism-design model of social-network auctions asks buyers not only to report their values, but also to invite their neighbors, despite the fact that doing so may create additional competition (Li et al., 2017). Subsequent work studies revenue objectives and structural limitations in such settings, including optimal diffusion auctions (Zhang et al., 2024), multi-unit extensions (Fang et al., 2023), combinatorial diffusion auctions (Li et al., 2024b), privacy-preserving diffusion incentives (Jia et al., 2023), and approximate revenue maximization for diffusion auctions (Huang et al., 2025).

Cost-aware diffusion mechanism. A closely related line of work studies one-shot auctions in intermediated economic networks where transaction costs are incurred when a selected trade passes through intermediaries. Li et al. (2018) introduce IDM-TC and the Customer Sharing Mechanism, which seek efficient, truthful, and weakly budget-balanced outcomes while inducing intermediaries to share information fully. Li et al. (2024a) subsequently extend this setting to multi-unit auctions and develop Critical Neighborhood Auctions to address the seller-deficit problem under transaction costs. More recently, Luo and Wang (2026) provide a decentralized implementation of transaction-cost diffusion auctions that preserves downstream-link privacy. In contrast, our setting models costs of information propagation itself: agents incur a cost when activating communication links, before the final allocation is known. We study repeated interaction under private stochastic valuations, where full propagation need not be optimal and agents learn selective, type-dependent diffusion policies from realized utilities. Relatedly, Zhang et al. (2023) study strategic connection control in a cost-sharing problem on directed acyclic graphs, where nodes may withhold outgoing edges to reduce their cost shares. Although their setting is not an auction or information-diffusion model, it similarly highlights the need for incentive design when network participants can strategically restrict connectivity under costs.

Learning-based auction design. A separate line of work studies how to learn auctions from data, usually in settings where the bidder set is exogenously fixed, and there is no strategic diffusion decision. From a statistical perspective, Syrgkanis (2017) studies sample complexity for learning optimal auctions, while Guo et al. (2021) considers robust learning of optimal auctions under corrupted samples. Neural approaches model allocation and payment rules as differentiable functions and optimize revenue subject to approximate incentive constraints: RegretNet (Duetting et al., 2019) initiated this approach, and later work improved the architecture and regret–revenue tradeoff using attention (Ivanov et al., 2022). Other extensions formulate auction learning as a two-player game with learned deviations (Rahme et al., 2021), incorporate preference and fairness constraints (Peri et al., 2021), use differentiable matching for richer allocation constraints (Curry et al., 2022), and

employ transformer architectures to solve equilibrium strategies across auction games (Huang et al., 2024).

Taken together, existing diffusion-auction work largely designs analytic mechanisms for networked participation, while learning-based auction design largely assumes a fixed participant set; to the best of our knowledge, prior work has not studied learning in network auctions where agents strategically decide whether, and to whom, to diffuse information over repeated selling interactions, potentially under diffusion costs.

Appendix B. Extended Preliminaries: Costless One-shot VCG

This appendix makes explicit the costless one-shot Network VCG primitives used as the baseline in Section 2. In particular, we define participation, critical diffusion nodes, the IDM allocation rule, and the resulting gross IDM payoff before any diffusion costs are introduced.

Participation and realized diffusion graph. Given a diffusion profile $\mathbf{x}$, let $\mathcal{G}_{\mathbf{x}} \subseteq \mathcal{G}$ denote the directed graph containing exactly the active information links. An agent j participates in the auction if and only if there exists a directed path of active information links from the seller s to j in $\mathcal{G}_{\mathbf{x}}$.

We denote the reachable set in Equation ($\mathcal{R}$) by

$$\mathcal{R}(\mathbf{x}) := \big\{ j \in \mathcal{N} \colon \text{there exists an active directed path from } s \text{ to } j \text{ in } \mathcal{G}_{\mathbf{x}} \big\}. \tag{$\mathcal{R}$}$$

Agents outside $\mathcal{R}(\mathbf{x})$ are uninformed and therefore do not submit bids.

Critical diffusion nodes. For a participating agent $j \in \mathcal{R}(\mathbf{x})$, another agent i is a *critical diffusion node* (CDE) if *every* active path from the seller s to j passes through i. Equation (d_i) defines the set of participating agents for whom i is a CDE:

$$d_i(\mathbf{x}) := \big\{ j \in \mathcal{R}(\mathbf{x}) \colon i \text{ is a critical diffusion node for } j \big\}. \tag{d_i}$$

By convention, $i \in d_i(\mathbf{x})$ whenever i participates.

For any subset $\mathcal{S} \subseteq \mathcal{N}$, Equation ($b^*_{-\mathcal{S}}$) defines the highest reachable bid outside $\mathcal{S}$:

$$b^*_{-\mathcal{S}}(\mathbf{x}) := \max_{j \in \mathcal{R}(\mathbf{x}) \setminus \mathcal{S}} b_j \,, \tag{$b^*_{-\mathcal{S}}$}$$

with the convention that the maximum is zero when the set is empty. Equation (m) defines the highest-bidding participating agent:

$$m \in \arg\max_{j \in \mathcal{R}(\mathbf{x})} b_j. \tag{m}$$

Ties are broken by a fixed deterministic rule.

Diffusion critical sequence. Equation ($\mathcal{K}_m$) defines the critical diffusion sequence:

$$\mathcal{K}_m(\mathbf{x}) = (\kappa_1, \kappa_2, \ldots, \kappa_L = m), \tag{$\mathcal{K}_m$}$$

the ordered sequence of critical diffusion nodes from the seller to the highest bidder m. Hence, for any index $\ell < L$, node κ_ℓ is a necessary antecedent for $\kappa_{\ell+1}$; thus, every active path from the seller to the highest bidder must pass through the sequence $\mathcal{K}_m(\mathbf{x})$ in the prescribed order.

VCG allocation rule. The auction mechanism adopts an efficient allocation policy π^*, which assigns the item to the participating buyer declaring the highest valuation. Therefore, the auction winner w coincides with the highest bidder of the active network, as stated in Equation (w):

$$w = m. \tag{w}$$

Given a feasible action profile $\mathbf{a} = (a_i)_{i \in \mathcal{N}}$, the realized social welfare of the allocation is:

$$W(\mathbf{a}) = \sum_{j \in \mathcal{N}, a_j \neq null} \pi_j^*(\mathbf{a}) b_j. \tag{W}$$

The VCG payment rule imposes a monetary debt on each participant equivalent to the marginal devaluation, or negative externality, caused by their presence on the rest of the network nodes. The formal payment of agent i, denoted as $p_i^{vcg}(\mathbf{a})$, is explicitly defined as the difference between the maximum social welfare theoretically obtainable by excluding the entire diffusion branch dependent on i (i.e. $W(\mathbf{a}_{-d_i(\mathbf{x})})$) and the actual collective welfare achieved by the other agents in the current allocation:

$$p_i^{vcg}(\mathbf{a}) = W(\mathbf{a}_{-d_i(\mathbf{x})}) - (W(\mathbf{a}) - \pi_i^*(\mathbf{a}) b_i)\,. \tag{p_i}$$

The net utility of agent i is equal to the true valuation of the allocation minus the required VCG payment: $u_i = \pi_i^*(\mathbf{a}) v_i - p_i^{vcg}(\mathbf{a})$. Substituting the explicit payment equation yields:

$$u_i^{vcg}(\mathbf{a}, (\pi^*, p^{vcg})) = W(\mathbf{a}) + \pi_j^*(\mathbf{a})(v_i - b_i) - W(\mathbf{a}_{-d_i(\mathbf{x})}). \tag{u_i^{vcg}}$$

Under the assumption of rational behavior, agents adopt the truthful revelation strategy (i.e., setting $b_i = v_i$), reducing the utility to the absolute marginal contribution of the agent to the overall social welfare of the system:

$$u_i^{vcg} = W(\mathbf{a}) - W(\mathbf{a}_{-d_i(\mathbf{x})}).$$

From this, three precise gross payoff configurations u_i^{vcg}

Definition 11 (gross VCG payoff). *Given bids and diffusion decisions* $(\mathbf{b}, \mathbf{x})$*, the gross payoff* $u_i^{idm}(\mathbf{b}, \mathbf{x})$*, for agent* i *with true valuation* v_i*, is defined as follows:*

- ***The Winner*** $(i = w)$***.*** *The winner receives the item* $(\pi_w^*(\mathbf{a}) = 1)$*, meaning the total social welfare is exactly their valuation:* $W(\mathbf{a}) = v_w$*. If node* w *had not participated, the item would have been assigned to the highest bidder outside its diffusion branch, generating a counterfactual welfare of* $W(\mathbf{a}_{-d_w}) = v_{-d_w}^*$*. The winner makes a payment equal to the second-highest bid external to their branch, obtaining a gross payoff of:*

$$u_i^{vcg} = v_w - b_{-d_w}^*.$$

- ***On-path Critical Diffusion Node*** $(i \in \mathcal{K}_w(\mathbf{x}) \setminus \{w\})$***.*** *These agents do not receive the item* $(\pi_i^*(\mathbf{a}) = 0)$*, but their diffusion action enabled the inclusion of the winner in the auction. The realized welfare is* $W(\mathbf{a}) = v_w$*. In the absence of* i*'s contribution, neither the winner nor anyone else in* d_i *would have participated, reducing the welfare to* $W(\mathbf{a}_{-d_i(\mathbf{x})}) = v^*_{-d_i(\mathbf{x})}$*. The VCG mechanism rewards them for this informational contribution:*

$$u_i^{vcg} = v_w - v^*_{-d_i(\mathbf{x})}.$$

- ***Non-Critical Agents.*** *For any agent* i *who is neither the winner nor an on-path critical node to the winner, removing their respective branch does not alter the efficient allocation* $W(\mathbf{a}) = W(\mathbf{a}_{-d_i(\mathbf{x})}) = v_w$*. Their marginal contribution is zero:*

$$u_i^{vcg} = 0.$$

Appendix C. Algorithmic Form of LDM and Stochastic-Gradient Derivation

This section provides the formal derivation for the gradient update rule used in Section 4 and gives the pseudocode of the resulting algorithm.

To maximize the ex-ante expected objective $J_i(\theta)$, rational agents employ gradient ascent. We compute the gradient of the objective function with respect to agent i's policy parameters, θ_i. By passing the gradient operator inside the expectation and summation, we isolate the policy of agent i, as the policies of competitors (σ_k for $k \neq i$) are independent of θ_i:

$$\nabla_{\theta_i} J_i(\theta) = \mathbb{E}_{\mathbf{v}\sim F}\left[\sum_{\mathbf{x}\in\mathbf{X}} u_i(\mathbf{v},\mathbf{x})\left(\prod_{k\neq i}\sigma_k(\mathbf{x}_k|v_k;\theta_k)\right)\nabla_{\theta_i}\sigma_i(\mathbf{x}_i|v_i;\theta_i)\right].$$

We apply the standard log-derivative identity, $\nabla f(z) = f(z)\nabla \log f(z)$, to agent i's policy:

$$\nabla_{\theta_i}\sigma_i(\mathbf{x}_i|v_i;\theta_i) = \sigma_i(\mathbf{x}_i|v_i;\theta_i)\nabla_{\theta_i}\log\sigma_i(\mathbf{x}_i|v_i;\theta_i).$$

Substituting this identity back into the gradient equation reconstructs the full joint probability measure $\prod_{k\in\mathcal{N}}\sigma_k$, allowing us to condense the expression into an expectation over both $\mathbf{v}$ and $\mathbf{x}$:

$$\nabla_{\theta_i} J_i(\theta) = \mathbb{E}_{\mathbf{v},\mathbf{x}}[u_i(\mathbf{v},\mathbf{x})\nabla_{\theta_i}\log\sigma_i(\mathbf{x}_i|v_i;\theta_i)].$$

Mathematical Derivation of the Log-Likelihood Gradient: To compute the gradient for a specific edge activation parameter $\theta_{i,j}$, we first take the natural logarithm of the joint policy:

$$\log\sigma_i(\mathbf{x}_i(t)\mid v_i(t);\theta_i) = \sum_{j\in r_i}\left[x_{i,j}(t)\log\mathbb{P}_{i,j}(\theta_{i,j}(v_i(t))) + (1-x_{i,j}(t))\left(1-\mathbb{P}_{i,j}(\theta_{i,j}(v_i(t)))\right)\right].$$

Taking the partial derivative with respect to $\theta_{i,j}$ requires applying the chain rule. We know the derivative of the logistic sigmoid function is $\frac{\partial\mathbb{P}_{i,j}}{\partial\theta_{i,j}} = \mathbb{P}_{i,j}(1-\mathbb{P}_{i,j})$. Therefore:

$$\begin{aligned}\frac{\partial\log\sigma_i}{\partial\theta_{i,j}} &= \frac{x_{i,j}}{\mathbb{P}_{i,j}}\mathbb{P}_{i,j}(1-\mathbb{P}_{i,j}) - \frac{1-x_{i,j}}{1-\mathbb{P}_{i,j}}\mathbb{P}_{i,j}(1-\mathbb{P}_{i,j}) = \\ &= x_{i,j}(1-\mathbb{P}_{i,j}) - (1-x_{i,j})\mathbb{P}_{i,j} = x_{i,j} - \mathbb{P}_{i,j}.\end{aligned}$$

Thus, the exact gradient component for parameter $\theta_{i,j}$ is:

$$\frac{\partial J_i(\theta)}{\partial\theta_{i,j}} = \mathbb{E}_{\mathbf{v},\mathbf{x}}\left[u_i(\mathbf{v},\mathbf{x})(x_{i,j}-\mathbb{P}_{i,j})\right].$$

Observation 1 (single-sample gradient estimator). *Calculating the exact expectation $\mathbb{E}_{\mathbf{v},\mathbf{x}}[\cdot]$ is strictly impossible for the agent because the joint measure depends on the unknown private types and policies of competitors $\mathbf{v}_{-i},\mathbf{x}_{-i}$. The fundamental approximation lies in replacing the intractable true expectation with a single-sample Monte Carlo estimate drawn from the environment at time step t. We define the stochastic gradient estimator $\hat{g}_{i,j}(t)$ as:*

$$\hat{g}_{i,j}(t) = u_i(t)(x_{i,j}(t) - \mathbb{P}_{i,j}(t)).$$

C.1. LDM Pseudocode

The following Algorithm 1 summarizes the resulting decentralized LDM update.

Algorithm 1: Learning Diffusion Mechanism (LDM)

Input: Network $G = (N, E)$, initially informed set S_0, cost $c > 0$, stepsizes $\{\alpha_t\}_{t\geq 1}$ Initialize diffusion-policy parameters θ_i for each agent $i \in N$ **for** $t = 1, 2, \ldots, T$ **do**

Draw private valuations $v_i(t) \sim F_i$ for all agents i Set informed set $R_t \leftarrow S_0$ **for** *each informed agent* $i \in R_t$ **do**

Bid truthfully: $b_i(t) \leftarrow v_i(t)$ **for** *each neighbor* $j \in r_i$ **do**

Compute $P_{i,j}(t) = P_{i,j}(v_i(t); \theta_{i,j}(t))$ Sample diffusion action $x_{i,j}(t) \sim \mathrm{Bernoulli}(P_{i,j}(t))$ **if** $x_{i,j}(t) = 1$ **then**

Add j to the informed set R_t

end

end

end

Run Network VCG on the realized reachable set R_t Each informed agent observes realized net utility $u_i(t) = u_i^{\mathrm{vcg}}(b(t), x(t)) - c\sum_{j\in r_i} x_{i,j}(t)$ **for** *each informed agent* $i \in R_t$ *and neighbor* $j \in r_i$ **do**

Update $\theta_{i,j}(t+1) = \theta_{i,j}(t) + \alpha_t u_i(t)\big(x_{i,j}(t) - P_{i,j}(t)\big)$

end

end

Appendix D. Technical Appendix: Missing Proofs of Main Part Statements

D.1. Proof of Proposition 1

Proof [Proof of Proposition 1] The agent's net utility $u_i(\mathbf{b}(t), \mathbf{x}(t))$ is additively separable into the gross payoff u_i^{vcg} and the total propagation cost $c\sum_{j\in r_i} x_{i,j}$. The cost component is strictly a function of the diffusion decision $\mathbf{x}_i(t)$ and is entirely independent of the chosen bid $b_i(t)$. Due to the intrinsic properties of the VCG design, the gross payoff u_i^{vcg} aligns individual utility with global social welfare, making $b_i(t) = v_i(t)$ a dominant strategy for any realized participant subgraph. Since the bid does not influence the cost term, the choice $b_i(t) = v_i(t)$ maximizes the total net utility. Therefore, truthful bidding weakly dominates any other bidding strategy. ■

D.2. Proof of Lemma 3

Proof [Proof of Lemma 3] Assuming truthful valuation reporting $(b_i = v_i)$ as established by Proposition 1, the net utility of agent i is additively separable into the gross VCG payoff and the localized propagation costs:

$$u_i(\mathbf{v}, \mathbf{x}) = u_i^{vcg}(\mathbf{v}, \mathbf{x}) - c\sum_{j\in r_i} x_{i,j}(t).$$

By the definition of the VCG payment rule, the gross payoff $u_i^{vcg}(\mathbf{v}, \mathbf{x})$ is the marginal difference between the realized social welfare, $W(\mathbf{v}, \mathbf{x})$, and the counterfactual social welfare obtained if i's downstream diffusion branch $d_i(\mathbf{x})$ were excluded:

$$u_i^{vcg}(\mathbf{v}, \mathbf{x}) = W(\mathbf{v}, \mathbf{x}) - W(\mathbf{v}, \mathbf{x}_{-d_i(\mathbf{x})}).$$

Substituting the gross payoff into the net utility function yields:

$$u_i(\mathbf{v}, \mathbf{x}) = W(\mathbf{v}, \mathbf{x}) - W(\mathbf{v}, \mathbf{x}_{-d_i(\mathbf{x})}) - c\sum_{j\in r_i} x_{i,j}.$$

We introduce a counterfactual potential function $\Phi_{-i}(\mathbf{v}, \mathbf{x}_{-i})$ defined as the maximum penalized social welfare achievable under the strict exclusion of agent i. This exclusion inherently truncates the propagation of information to branch $d_i(\mathbf{x})$, effectively reducing the gross social welfare to $W(\mathbf{v}, \mathbf{x}_{-d_i(\mathbf{x})})$ and eliminating agent i's localized routing costs from the system's aggregate cost:

$$\Phi_{-i}(\mathbf{v}, \mathbf{x}_{-i}) = W(\mathbf{v}, \mathbf{x}_{-d_i(\mathbf{x})}) - c\sum_{k\in\mathcal{N}\setminus\{i\}}\sum_{l\in r_k} x_{k,l}.$$

Evaluating the difference between the global potential $\Phi(\mathbf{v}, \mathbf{x})$ and the counterfactual potential $\Phi_{-i}(\mathbf{v}, \mathbf{x}_{-i})$ allows us to isolate the welfare and network cost differentials:

$$\Phi(\mathbf{v}, \mathbf{x}) - \Phi_{-i}(\mathbf{v}, \mathbf{x}_{-i}) = \left[W(\mathbf{v}, \mathbf{x}) - W(\mathbf{v}, \mathbf{x}_{-d_i(\mathbf{x})})\right] - c\left[\sum_{k\in\mathcal{N}}\sum_{l\in r_k} x_{k,l} - \sum_{k\in\mathcal{N}\setminus\{i\}}\sum_{l\in r_k} x_{k,l}\right].$$

The algebraic difference in aggregate network costs simplifies directly to the local communication costs sustained exclusively by agent i:

$$\sum_{k\in\mathcal{N}}\sum_{l\in r_k} x_{k,l} - \sum_{k\in\mathcal{N}\setminus\{i\}}\sum_{l\in r_k} x_{k,l} = \sum_{j\in r_i} x_{i,j}.$$

Consequently, the potential difference reduces to:

$$\Phi(\mathbf{v},\mathbf{x}) - \Phi_{-i}(\mathbf{v},\mathbf{x}_{-i}) = W(\mathbf{v},\mathbf{x}) - W(\mathbf{v},\mathbf{x}_{-d_i(\mathbf{x})}) - c\sum_{j\in r_i} x_{i,j}.$$

This resulting expression is identically equal to the agent's net utility $u_i(\mathbf{v},\mathbf{x})$, thereby concluding the formal verification of the identity. ■

D.3. Proof of Theorem 4

Proof [Proof of Theorem 4] Let $\mathbf{x}_i$ represent the initial diffusion strategy of agent i, and assume a unilateral deviation to an alternative strategy $\mathbf{x}'_i$. The joint action profile of all other agents, $\mathbf{x}_{-i}$, remains strictly constant.

To establish an exact potential game, we must demonstrate that the change in the individual's net utility exactly equals the change in the global potential function.

From Lemma 3, we know the agent's utility is equivalent to their marginal contribution to the potential function. Therefore, the individual utility differential caused by the deviation is expressed as:

$$\begin{aligned}\Delta u_i &= u_i(\mathbf{v},(\mathbf{x}'_i,\mathbf{x}_{-i})) - u_i(\mathbf{v},(\mathbf{x}_i,\mathbf{x}_{-i})) = \\ &= \left[\Phi(\mathbf{v},(\mathbf{x}'_i,\mathbf{x}_{-i})) - \Phi_{-i}(\mathbf{v},\mathbf{x}_{-i})\right] - \left[\Phi(\mathbf{v},(\mathbf{x}_i,\mathbf{x}_{-i})) - \Phi_{-i}(\mathbf{v},\mathbf{x}_{-i})\right].\end{aligned}$$

Crucially, because the counterfactual social welfare $\Phi_{-i}(\mathbf{v},\mathbf{x}_{-i})$ is mathematically defined upon the exclusion of node i, it is entirely independent of the specific diffusion action ($\mathbf{x}_i$ or $\mathbf{x}'_i$) taken by agent i. Consequently, this term cancels out of the algebraic computation:

$$\Delta u_i = \Phi(\mathbf{v},(\mathbf{x}'_i,\mathbf{x}_{-i})) - \Phi(\mathbf{v},(\mathbf{x}_i,\mathbf{x}_{-i})) = \Delta\Phi.$$

The unilateral change in individual utility equals the exact unilateral change in the global Social Welfare function, satisfying the formal mathematical conditions of an exact potential game. ■

D.4. Proof of Theorem 6

Proof [Proof of Theorem 6] By Proposition 1, truthful bidding is a weakly dominant strategy regardless of the diffusion decision. Consequently, strategic interaction reduces to the diffusion game

alone, and the mechanism is strategically equivalent to a Bayesian game in which each agent $i \in \mathcal{N}$ observes a private valuation $v_i \in [0, v_{\max}]$ and chooses a diffusion action $\mathbf{x}_i \in \mathbf{X}_i = \{0,1\}^{|r_i|}$. Since the social network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is finite, every neighborhood r_i is finite, implying that each action set $\mathbf{X}_i$ is finite. A behavioral strategy of agent i is a measurable mapping

$$\sigma_i \ : \ [0, v_{\max}] \to \Delta(\mathbf{X}_i),$$

where $\Delta(\mathbf{X}_i)$ denotes the simplex of probability measures over the finite action set $\mathbf{X}_i$. Hence, mixed strategies are well defined for every realization of the private type. For every valuation profile $\mathbf{v}$ and action profile $\mathbf{x}$, the realized utility is $u_i(\mathbf{v}, \mathbf{x}) = u_i^{vcg}(\mathbf{v}, \mathbf{x}) - c \sum_{j \in r_i} x_{i,j}(t)$. Since social welfare is bounded by the largest attainable valuation,

$$0 \leq W(\mathbf{v}, \mathbf{x}) \leq v_{\max},$$

the VCG payoff is uniformly bounded. Moreover,

$$0 \leq c \sum_{j \in r_i} x_{i,j}(t) \leq c|r_i|$$

which yields

$$|u_i(\mathbf{v}, \mathbf{x})| \leq v_{\max} + c|r_i|$$

Therefore the utility function is bounded and integrable with respect to the product distribution of private valuations. The only possible discontinuities of u_i arise when two or more agents have identical valuations, because in this case the identity of the auction winner may depend on the tie-breaking rule. However, each valuation distribution F_i is assumed to be absolutely continuous with respect to the Lebesgue measure on $[0, v_{\max}]$. Consequently, $\mathbb{P}(v_i = v_j) = 0$ for all $i \neq j$, so the set of discontinuity points has product measure zero. Hence the expected utility $J_i(\sigma) = \mathbb{E}_{\mathbf{v},\mathbf{x}}\left[u_i(\mathbf{v}, \mathbf{x})\right]$ is well defined and continuous with respect to the mixed strategy profile.

Since every action space $\mathbf{X}_i$ is finite and every type space is compact, the Bayesian diffusion game satisfies the standard hypotheses of the existence theorem (Milgrom and Weber (1985)) for Bayesian games with continuous type spaces and finite action sets. Therefore there exists a measurable behavioral strategy profile $\sigma^* = (\sigma_1^*, \cdots, \sigma_n^*)$, such that, for every agent i,

$$J_i(\sigma_i^*, \sigma_{-i}^*) \geq J_i(\sigma_i, \sigma_{-i}^*)$$

for every admissible behavioral strategy σ_i.

Finally, since the valuation distributions are stationary and strategies depend only upon the current realization of the private type, the equilibrium strategy profile is independent of time. Hence σ^* is a stationary Bayes–Nash equilibrium of the Learning Diffusion Mechanism. ■

D.5. Proof of Proposition 8

Proof [Proof of Proposition 8] We explicitly define the noise term as:

$$M_{i,j}(t+1) = \hat{g}_{i,j}(t) - \frac{\partial J_i(\theta(t))}{\partial \theta_{i,j}} = \\ = u_i(t)(x_{i,j}(t) - \mathbb{P}_{i,j}(t)) - \mathbb{E}_{\mathbf{v},\mathbf{x}}\left[u_i(\mathbf{v},\mathbf{x})(x_{i,j} - \mathbb{P}_{i,j})\right].$$

Taking the conditional expectation of the noise with respect to the filtration $\mathcal{F}_t$:

$$\mathbb{E}\left[M_{i,j}(t+1) \mid \mathcal{F}_t\right] = \mathbb{E}\left[\hat{g}_{i,j}(t) \mid \mathcal{F}_t\right] - \mathbb{E}\left[\frac{\partial J_i(\theta(t))}{\partial \theta_{i,j}} \mid \mathcal{F}_t\right].$$

By definition, the parameters $\theta(t)$ are completely $\mathcal{F}_t$-measurable (deterministic given the history). The conditional expectation of the stochastic estimator over the sampling distributions of $\mathbf{v}$ and $\mathbf{x}$ recovers the exact analytical gradient $\frac{\partial J_i(\theta(t))}{\partial \theta_{i,j}}$. Thus:

$$\mathbb{E}\left[M_{i,j}(t+1) \mid \mathcal{F}_t\right] = \frac{\partial J_i(\theta(t))}{\partial \theta_{i,j}} - \frac{\partial J_i(\theta(t))}{\partial \theta_{i,j}} = 0.$$

This guarantees that the stochastic approximation introduces zero bias in expectation. ■

D.6. Proof of Theorem 10

Proof [Proof of Theorem 10] Under the continuous-time approximation of the REINFORCE algorithm, the trajectory of agent i's policy parameters follows the exact gradient of their ex-ante expected utility:

$$\frac{d\theta_i}{dt} = \nabla_{\theta_i} J_i(\theta).$$

By Lemma 3, for any ex-post realization of valuations $\mathbf{v}$ and actions $\mathbf{x}$, the net utility satisfies the exact potential identity: $u_i(\mathbf{v},\mathbf{x}) = \Phi(\mathbf{v},\mathbf{x}) - \Phi_{-i}(\mathbf{v},\mathbf{x}_{-i})$. Because the expectation operator is linear, taking the ex-ante expectation of both sides perfectly preserves this structural identity:

$$J_i(\theta) = \Phi(\theta) - \Phi_{-i}(\theta_{-i}).$$

Because the counterfactual potential $\Phi_{-i}(\theta_{-i})$ is completely independent of agent i's strategic parameters θ_i. Therefore, computing the partial gradient with respect to θ_i eliminates the counterfactual term, aligning the individual expected utility gradient exactly with the global expected potential gradient:

$$\nabla_{\theta_i} J_i(\theta) = \nabla_{\theta_i} \Phi(\theta).$$

Substituting this identity into the continuous-time behavioral ODE reveals that the independent multi-agent learning dynamics collapse into a centralized pseudo-gradient ascent on the expected potential function:

$$\frac{d\theta}{dt} = \nabla \Phi(\theta).$$

To establish global asymptotic stability, we designate the expected Social Welfare $\Phi(\theta)$ as our candidate Lyapunov function $\mathcal{L}(\theta)$. The time evolution of the Lyapunov function along the system's trajectory is determined via the multivariate chain rule:

$$\frac{d\mathcal{L}(\theta(t))}{dt} = \left\langle \nabla\Phi(\theta)), \frac{d\theta}{dt} \right\rangle = \sum_{i\in\mathcal{N}} \left(\nabla_{\theta_i}\Phi(\theta)\right)^T \left(\nabla_{\theta_i}\Phi(\theta)\right) = \|\nabla\Phi(\theta(t))\|^2 \geq 0.$$

This establishes that the expected global Social Welfare monotonically increases over time. Because all private valuations are strictly bounded from above by $v_{\max}$, the expected global potential $\Phi(\theta)$ is similarly bounded above for all θ. Since $\Phi(\theta)$ is bounded above and monotonically increasing along the solution trajectories, the limit

$$\lim_{t\to\infty} \Phi(\theta(t)),$$

exists and is finite. Because the system constitutes a pure gradient flow on a smooth, bounded potential manifold, it directly follows from standard dynamical systems theory (via Barbalat's Lemma) that the gradient must asymptotically vanish:

$$\lim_{t\to\infty} \|\nabla\Phi(\theta(t))\| = 0.$$

Consequently, the continuous-time strategy dynamics are mathematically guaranteed to converge to a stationary point of $\Phi(\theta)$, corresponding to a Bayes-Nash Equilibrium of the learning diffusion game. ■

Appendix E. A Case Study of Example Information Diffusion Network under Varying Costs

Through the following example, we show how local propagation costs alter agents' diffusion incentives and progressively reduce the size of the realized auction network. The objective is not to provide a complete equilibrium characterization of the induced diffusion game, but rather to demonstrate the incentive forces that the Learning Diffusion Mechanism (LDM) is designed to induce.

Consider the global (unrealized) diffusion network $\mathcal{G}$ depicted in Figure 2. The seller (s) is directly connected to three agents A, B and C with their respective values:

$$v_A = 1, \quad v_B = 3, \quad \text{and} \quad v_C = 2.$$

The remaining network is structured as follows:

- **Sub-network A:** Agent A connects to D ($v_D = 5$), who acts as a secondary hub informing F ($v_F = 10$) and G ($v_G = 8$). Agent G informs K ($v_K = 6$).
- **Sub-network B:** Agent B connects to E ($v_E = 7$).
- **Sub-network C:** Agent C informs E, J ($v_J = 4$), H ($v_H = 11$) and I ($v_I = 12$). Agent I serves as the sole bridge to the network's highest valuation node, L ($v_L = 13$).

In the following, assume truthful bidding, as Proposition 1 implies.

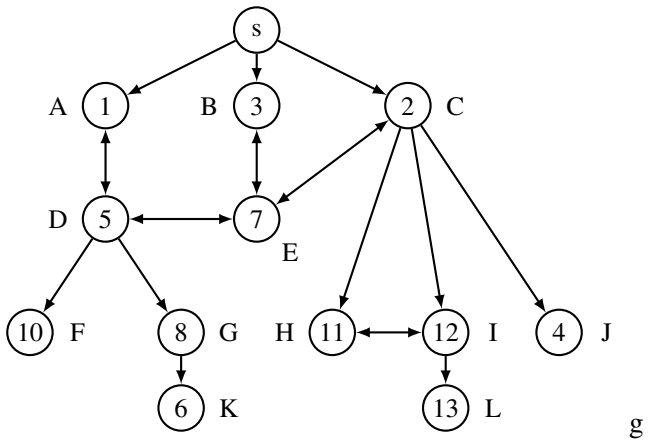


Figure 2: **Example of information diffusion network flow**, studied in Appendix E (unrealized graph). Each potential information propagation—that is, each edge in the above graph—has a cost c. The letters denote the node enumeration, whereas the numbers are the respective nodes' estimated value, coinciding with their bids.

The baseline: Costless Information Diffusion $(c = 0)$. When diffusion is costless $(c = 0)$, full network propagation is a weakly dominant strategy, and all edges are active. The highest valuation in the network is $v_L = 13$, so Agent L receives the item. The critical diffusion sequence leading to the winner is $\mathcal{K}_L = \{C, I, L\}$. The resulting VCG payoffs are as follows:

- **Agent L (Winner):** Removing L's diffusion branch leaves I as the highest remaining valuation. Therefore, $u_L^{vcg} = 13 - 12 = 1$.
- **Agent I (On-path Critical Node):** If agents I and L are excluded, the highest remaining valuation is $v_H = 11$. Hence $u_I^{vcg} = 13 - 11 = 2$.
- **Agent C (On-path Critical Node):** Removing the branch $d_C(\mathbf{x}) = \{C, J, H, I, L\}$, leaves F as the highest remaining valuation. Therefore $u_C^{vcg} = 13 - 10 = 3$.
- **Off-path Agents** All remaining agents are off the critical path leading to the winner and therefore receive zero VCG payoff.

Consequently, the efficient allocation is achieved and the winner is the globally highest-valued buyer.

The Repeated Game: Costly Diffusion $(c > 0)$.
When a marginal cost $c > 0$ is introduced per activated link, the deterministic nature of the valuations causes the collapse of "option value". Because non-critical agents face a strict mathematical guarantee of $u_i^{vcg} = 0$, any edge activation yields a strictly negative utility. Agents no longer diffuse to discover potentially high-value buyers; they optimize a deterministic tradeoff, leading to immediate and severe network pruning. To understand the resulting incentives, consider a sequence of increasingly restrictive diffusion profiles.

Incentive of agent I. Assume that:

- Agent C informs I;
- Agent A does not activate its branch;
- Agent C does not activate links E, J or H.

Under this profile, the active network contains only $s \to C \to I$, and potentially L.

- **If agent I informs L:** Agent L wins. Excluding the branch $\{I, L\}$ leaves Agent B as the highest active bidder. Thus $U_I(x_{I,L} = 1) = (13 - 3) - c = 10 - c$.
- **If agent I excludes L:** Agent I becomes the winner. The second highest active valuation is again $v_B = 3$, implying $U_I(x_{I,L} = 0) = 12 - 3 = 9$.

Hence, I prefers informing L whenever $10 - c > 9 \Longrightarrow c < 1$.

Incentive of agent C. Suppose that the branch of agent A is active, so that agent F ($v_F = 10$) participates. If I is informed but L is excluded, then I becomes the winner with value 12. Agent C's marginal contribution is then evaluated relative to the best alternative bidder outside its diffusion branch, namely F. Hence $u_C^{vcg} = 12 - 10 = 2$. If informing I requires activating a single link, agent C's net payoff becomes $u_C = 2 - c$. Accordingly, diffusion toward I is individually profitable whenever $c < 2$.

Incentive of agent A. Now consider agent A. If the branch leading to F is activated, the highest reachable valuation in that branch is $v_F = 10$. When the competing branch through C is inactive, F becomes the winner and the relevant competing valuation is $v_B = 3$. The resulting social welfare contribution equals $10 - 3 = 7$. Since activating the chain requires at least one costly diffusion action, the profitability of maintaining this breach decreases as c increases. In particular, once diffusion cost becomes comparable to the maximum attainable informational surplus generated by the branch, activation is no longer individually attractive.

Progressive Network Contraction. The example highlights a sequence of qualitative transitions.

- **Regime 1: The Monopolistic Broker** ($0 < c < 1$) The informational reward obtained by intermediaries exceeds the propagation cost. The path toward L remains active and the globally efficient allocation is achieved.
- **The Endogenous Winner and Unique Dominance** ($1 \leq c < 2$) Agent I prefers not to reveal information to L. The highest valued buyer is therefore excluded from the auction, and agent I becomes the winner.
- **Substitutes and the Coordination Game** ($2 \leq c < 7$) A structural bifurcation occurs at $c \geq 2$, Agent C can no longer profitably operate against an active Sub-network A, as its net utility would drop to $2 - c \leq 0$. The diffusion decisions of Agents A and C become perfect strategic substitutes, fracturing the mechanism into two mutually exclusive coordination equilibria:

- **Equilibrium** 1 **(C-Dominance).** Agent C activates the link to I. Observing this dominance, Agent A truncates its link to D. Agent F is never informed. Winner: Agent I.
- **Equilibrium** 2 **(A-Dominance).** Agent A pays c to inform D, and D pays c to inform F. Agent C, facing the external baseline $v_F = 10$, realizes diffusion is unprofitable ($12 - 10 - c \leq 0$) and truncates the link to I. Agent I and Agent L are never informed. Winner: Agent F.

- **Terminal Network Collapse ($c \geq 9$).** The cost of propagation exceeds the maximum possible critical node reward. Agents A, B, and C strictly truncate all outgoing links. The diffusion process terminates immediately at the source, restricting the auction to the seller's immediate, uninformed neighborhood. The item is allocated to Agent B ($v_B = 3$), generating severe allocative inefficiency.

This example illustrates the central mechanism underlying LDM. In the classical diffusion-auction literature, full propagation is typically incentive compatible because communication is assumed to be free. Once even small propagation costs are introduced, agents begin to compare the informational value generated by a diffusion link against its direct cost. As a result, efficient information flow is no longer guaranteed, and the realized auction network becomes an endogenous outcome of strategic behavior. The learning dynamics introduced in Section 4 can be interpreted as a decentralized procedure through which agents gradually discover these profitable and unprofitable diffusion patterns without requiring knowledge of the global network topology or the valuation distributions of other participants.

Equilibrium Degeneration The existence of a stationary Bayes-Nash Equilibrium relies on Glicksberg's Fixed-Point Theorem (Glicksberg, 1952), which assumes continuity of the expected utility function, condition satisfied when valuation distributions F_i are absolutely continuous. If valuations v_i are fixed these distribution collapse to point mass. In the presence of identical valuation, the deterministic tie-breaking rules create discontinuities in the payoff functions, which invalidates the existence proof. To address this, we employ two alternative formulations:

- **Finite Strategic Game:** with constant valuations, the model reduces from Bayesian game to standard game with complete information. The strategy profile space simplifies to a finite mixed strategy over action profiles $\mathbf{X}_i \in \{0,1\}^{|r_i|}$. Existence of mixed-strategy Nash Equilibrium is guaranteed by the finiteness of agent set $\mathcal{N}$ and the action spaces.
- **Perturbation Method:** To preserve the Bayesian framework, we apply an ϵ-perturbation to the deterministic valuations,
$$v_i \approx \mathcal{U}(v_i^* - \epsilon, v_i^* + \epsilon).$$
This ensures the absolute continuity of F_i for any $\epsilon > 0$, thereby smoothing the expected utility function and restoring the existence of a stationary Bayes-Nash Equilibrium.

Appendix F. Simulated LDM Dynamics for the Example in Appendix E

To illustrate the learning dynamics predicted by the theory, we run a numerical simulation. The computational experiment is implemented on the directed social network topology $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ in Figure 2, studied in Appendix E. To ensure the absolute continuity of the expected utility functions and prevent payoff discontinuities arising from deterministic tie-breaking rules, agent valuations are sampled from a uniform distribution subject to infinitesimal perturbations. We impose a strictly positive marginal propagation cost $c = 0.1$, which induces agents to strategically prune the network to avoid unnecessary propagation costs.

Throughout the simulation, each strategic agent $i \in \mathcal{N}$ iteratively refines its parameterized diffusion policy $\theta_{i,j}$ employing a REINFORCE-based stochastic gradient ascent mechanism. The discrete-time recursive update rule governing the evolution of the edge-activation propensities:

$$\theta_{i,j}(t+1) = \theta_{i,j}(t) + \alpha_t u_i(t)(x_{i,j}(t) - \mathbb{P}_{i,j}(t)),$$

where the learning rate is uniformly set to $\alpha_t = 0.01$.

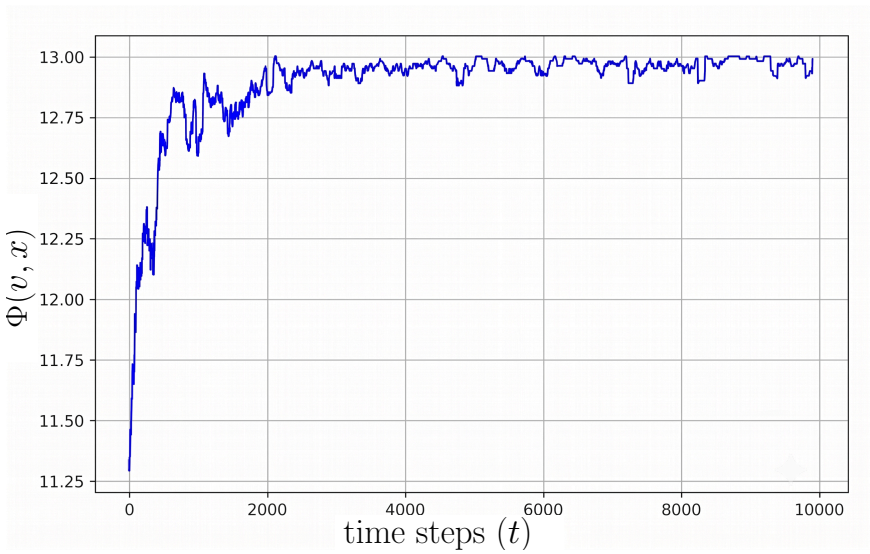
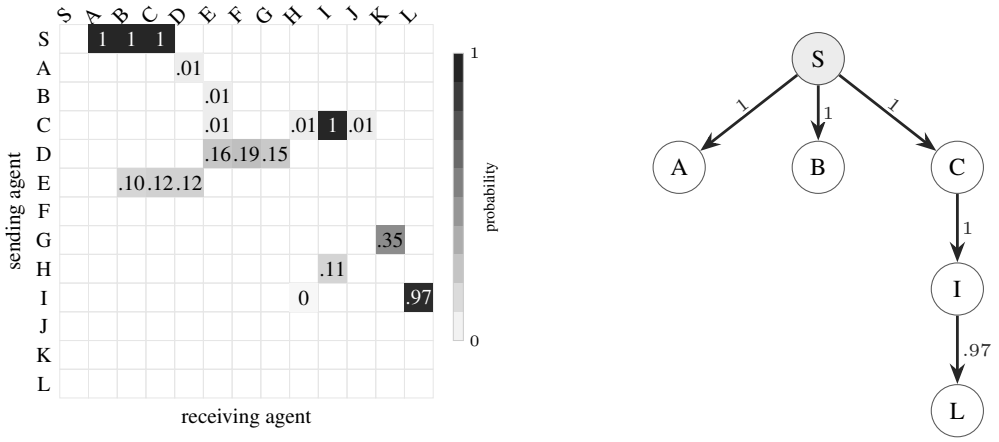


Figure 3: **Mean-field learning improves welfare under costly diffusion.** Expected social welfare over $T = 10000$ REINFORCE updates on the network in Figure 2, with propagation cost $c = 0.1$ and learning rate $\alpha_t = 0.01$. Independent policy-gradient dynamics increase welfare as agents learn selective diffusion policies that trade off VCG rewards against communication costs.

The global convergence trajectory of the multi-agent system, evaluated over a discrete time horizon of $T = 10000$ steps, is represented in Figure 3. The empirical dynamics illustrate the system's capacity to asymptotically optimize the trade-off between the extraction of VCG-based critical-node rewards and the accumulation of exogenous propagation penalties.



(a) Converged activation matrix.

(b) Thresholded learned diffusion graph.

Figure 4: **Converged LDM diffusion policy**, using $c = 0.1$, after 10K steps. Left: grayscale activation matrix, where darker cells indicate larger probabilities from sending agent (row) to receiving agent (column). Right: learned active graph after thresholding at .5, shown edges have activation probability $\geq .5$.

Upon reaching convergence, the terminal diffusion matrix, visualized as a heatmap of converged network activation probabilities in Figure 4, demonstrates near-deterministic edge activation strictly along the network's maximal-value critical paths. Specifically, Figure 4 highlights the isolation of optimal routing sequences, such as $S \to \{A, B, C\}$ and the subsequent path $C \to I \to L$. This resultant topological configuration empirically corroborates the learned strategic behavior and theoretical equilibrium characterized in Regime 1 ($0 < c < 1$) of the example in Appendix E. This simulation confirms that decentralized policy gradient updates successfully guide the strategic network toward a highly selective, welfare-maximizing active subgraph.

Appendix G. The Information Diffusion Mechanism (IDM).

To contextualize the comparative advantages of the Network VCG mechanism utilized in this work, we formalize the allocation and payment rules of the Information Diffusion Mechanism (IDM), which serves as the standard baseline in costless diffusion literature. We leverage the topological primitives, such as the critical diffusion sequence $\mathcal{K}_m(\mathbf{x}) = (\kappa_1, \kappa_2, \ldots, \kappa_L = m)$ and the highest external bid $b^*_{-\mathcal{S}}(\mathbf{x}) = \max_{j \in \mathcal{R}(\mathbf{x}) \setminus \mathcal{S}} b_j$, already defined in Appendix B.

IDM Allocation Rule. Unlike VCG, which strictly allocates to the globally highest bidder, IDM employs a sequential threshold condition along the critical diffusion sequence. The mechanism evaluates each critical node κ_l (for $l < L$) against the best alternative bid available outside the set of agents who depend on the subsequent critical node κ_{l+1} to participate.

The winner w is identified as the first agent κ_{l^*} in the sequence whose reported bid matches or exceeds the highest competing bid from agents outside the influence of the subsequent node in the sequence. Formally, the winner is

$$w = \kappa_{\ell^\star}, \qquad \ell^* = \min \left\{ \{\ell \in \{1, \ldots, L-1\} \mid b_{\kappa_\ell} \geq b^*_{-d_{\kappa_{\ell+1}}(\mathbf{x})}\} \cup L \right\} . \tag{1}$$

If no intermediate node satisfies the threshold condition, the item defaults to the globally highest bidder, $\kappa_L = m$.

Gross IDM Payoff. Given bids and diffusion decisions $(\mathbf{b}, \mathbf{x})$, the gross IDM payoff $u_i^{idm}(\mathbf{b}, \mathbf{x})$, for agent i with true valuation v_i, is distributed in a piecewise manner based on their position relative to the winner:

- **The Winner($i = w$):** The winning agent receives the item and pays the highest competing bid outside of their own downstream influence branch:

$$u_i^{idm} = v_i - b^*_{-d_i(\mathbf{x})} \, .$$

- **On path critical diffusion node ($i \in \mathcal{K}_w(\mathbf{x}) \setminus \{w\}$):** Agents who reside on the critical sequence prior to the winner do not receive the item, but are compensated for expanding the market. Their reward is exactly their marginal contribution to the competitive environment:

$$u_i^{idm} = b^*_{-d_{i+1}(\mathbf{x})} - b^*_{-d_i(\mathbf{x})} \, .$$

- **Non-Critical Agents:** Any agent who is neither the winner nor an on-path critical node leading to the winner receives a gross payoff of strictly zero:

$$u_i^{idm} = 0 \, .$$

On the Mathematical Failure of IDM to Induce an Exact Potential Game. To establish an Exact Potential Game under independent policy gradient dynamics, the unilateral change in an agent's

expected utility must identically match the unilateral change in a global scalar potential function. Lemma 3 demonstrates that the Network VCG mechanism achieves this by defining payoffs as the marginal contribution to the global social welfare $\Phi(\mathbf{v}, \mathbf{x})$, serving as the exact potential function. Conversely, the Information Diffusion Mechanism (IDM) breaks this identity due to a structural decoupling between its sequential reward allocations and the global welfare objective. This failure manifests through two primary structural limitations described next.

- **Severance from Global Welfare Maximization.** VCG guarantees that the item is allocated to the bidder with the maximum valuation in the reachable network, $\mathcal{R}(\mathbf{x})$. Consequently, any unilateral deviation by agent i that expands the network to discover a higher-valued buyer increases both $W(\mathbf{v}, \mathbf{x})$ and agent i's utility u_i^{vcg}. In contrast, IDM determines the winner $w = \kappa_{\ell^\star}$ via a sequential threshold rule along the critical path $\mathcal{K}_m(\mathbf{x})$, allocating the item to the first intermediary node whose bid satisfies $b_{\kappa_\ell} \geq b^*_{-d_{\kappa_{\ell+1}}(\mathbf{x})}$. Because IDM can allocate the item to an intermediate node κ_l even when a strictly higher valuation exists downstream (i.e., $v_{\kappa_\ell} \leq v_m$), the realized welfare $W_{idm}(\mathbf{v}, \mathbf{x})$, is generally less than the theoretical maximum $W(\mathbf{v}, \mathbf{x})$. Therefore, a unilateral diffusion action by agent i might shift IDM payment thresholds and winners, triggering discontinuous jumps in realized welfare that are not proportionally reflected in the agent's payoff.

- **Asymmetric and Localized Utility Gradients.** For an exact potential game to exist, a any unilateral deviation $\mathbf{x}_i \to \mathbf{x}'_i$, the change in gross utility (ignoring propagation costs for isolation) must satisfy:

$$\Delta u_i^{idm} = u_i(\mathbf{v}, (\mathbf{x}'_i, \mathbf{x}_{-i})) - u_i(\mathbf{v}, (\mathbf{x}_i, \mathbf{x}_{-i})) = W(\mathbf{v}, \mathbf{x}') - W(\mathbf{v}, \mathbf{x}) = \Delta W.$$

  Consider an on-path critical node i in diffusion profile $\mathbf{x}$, whose IDM rewards is $u_i^{idm}(\mathbf{x}) = b_{-d_{i+1}(\mathbf{x})} - b_{-d_i(\mathbf{x})}$. Suppose agent i unilaterally truncates the branch leading to $i+1$ such that i becomes the new winner. The new gross utility is:

$$u_i^{idm}(\mathbf{x}') = v_i - b^*_{-d_i(\mathbf{x}')}.$$

  Because the external branches not dependent on $i+1$ remain unchanged, $b^*_{-d_i(\mathbf{x}')} = b^*_{-d_i(\mathbf{x})}$ The deviation in agent i's IDM utility is:

$$\Delta u_i^{idm} = \left(v_i - b^*_{-d_i(\mathbf{x}')}\right) - \left(b_{-d_{i+1}(\mathbf{x})} - b_{-d_i(\mathbf{x})}\right) = v_i - b^*_{-d_{i+1}(\mathbf{x}')}.$$

  However, the true change in global welfare caused by this truncation is $\Delta W = v_i - v_w$ (where $v_w > v_i$ is the previous winner). As $b^*_{-d_{i+1}(\mathbf{x})} \neq v_w$ in general, the identity $\Delta u_i^{idm} = \Delta W$ is violated. The IDM's payments structure is thus unable to align individual incentives with global welfare maximization.