

Abstract citation ID: bbaf631.016

CELLSIGFLOW: A FLOW-BASED OPTIMIZATION FRAMEWORK FOR REALISTIC SINGLE-CELL AND SPATIAL INFERENCE OF INTERCELLULAR COMMUNICATION

Menghan Wang[‡], Junya Yang[‡], Jiaxing Chen^{*}

Guangdong Provincial/Zhuhai Key Laboratory IRADS, Beijing Normal-Hong Kong Baptist University, JinTong Road, 519087, Zhuhai, China

^{*} Corresponding author.

[‡] Equal Contribution.

Background: Cell-cell communication (CCC) is fundamental for tissue organization, developmental programs, and disease progression. The advent of single-cell and spatial transcriptomics has made it possible to infer ligand-receptor (L-R) interactions at unprecedented resolution, and numerous computational frameworks have been proposed to integrate gene expression with spatial information to predict putative signaling events.

Early approaches such as CellPhoneDB and CellChat (1) infer L-R interactions primarily at the cluster level. While these methods capture global signaling patterns, they cannot resolve communication at the level of individual cells. Moreover, they typically analyze each L-R pair in isolation, which limits their ability to reveal the complexity of real signaling networks that involve multiple ligands and receptors acting in concert.

Recent tools designed for spatial transcriptomics, including Giotto, NICHES, and COMMOT(2), incorporate neighborhood information or optimal transport to achieve cell-level predictions.

Despite these advances, there are still important limitations. Existing methods often overlook competition among receptors and ligands within a cell, and they rarely model the capacity constraints imposed by the ligand or receptor abundance. As a result, predicted communication edges may lack biological realism, showing either spurious absence or unrealistic saturation of signaling. Furthermore, most frameworks remain restricted to single-step L-R pairs, leaving multistep relay cascades largely unexplored.

To address these challenges, we propose a method called CellSigFlow, which formulates CCC as a bounded minimum-cost flow problem. In this framework, each cell is represented as a node and receptor-centric directed edges are constructed, with communication strength modeled as edge flow. Spatial distance and molecular affinity are incorporated into the cost function to penalize long-range interactions and favor biologically plausible binding. By introducing upper bounds proportional to receptor abundance and lower bounds informed by prior knowledge, the model enforces non-zero yet biologically constrained fluxes. This design enables the framework to capture competition among multiple receptors for the same ligand within a cell, resolve signaling at the single-cell level, and naturally integrate multiple L-R pairs to reconstruct complex communication networks beyond isolated interactions.

Methods: According to our method CellSigFlow, we formulate CCC as a bounded minimum-cost flow problem on a spatially constrained graph. Each cell is represented as a node. For a target cell j , and for each receptor R expressed in j , we create a directed edge (i,j,R) from a source cell i if the spatial distance d_{ij} is below a cutoff D , and there exists at least one ligand L expressed in i that can bind R in j . Thus, a pair of cells may be connected by multiple edges only to the extent of distinct receptors on the target; communication strength is modeled as the flow $f_{ij}^{(R)}$ on edge (i,j,R) .

For each receptor-centric edge, the flow is constrained by a lower bound $L_{ij}^{(R)}$ and an upper bound $U_{ij}^{(R)}$:

$$L_{ij}^{(R)} \leq f_{ij}^{(R)} \leq U_{ij}^{(R)}.$$

Upper bounds are proportional to receptor abundance in j together with the aggregate availability of ligands in i that bind R , while lower bounds incorporate prior knowledge (e.g., CellPhoneDB, CellChat), ensuring nonzero yet biologically plausible signaling at the receptor level. The complete bounded minimum-cost flow formulation requires that, in addition to satisfying upper and lower capacity constraints, the system obeys flow conservation at each node and minimizes the total transportation cost. The receptor-specific cost on edge is defined as

$$c_{ij}^{(R)} = \lambda_d d_{ij} + \lambda_a \psi(\text{affinity}_{ij}^{(R)}),$$

where d_{ij} denotes the spatial distance between cells and $\psi(\cdot)$ encodes a receptor-centric affinity term that aggregates the

contributions of all ligands in i capable of binding receptor R in j . The optimization objective is therefore to minimize the total cost

$$\min_f \sum_{(i,j,R)} c_{ij}^{(R)} f_{ij}^{(R)}$$

subject to

$$\sum_{(u,R)} L_{uj}^{(R)} \leq f_{uj}^{(R)} \leq U_{uj}^{(R)},$$

$$\sum_{(u,R)} f_{uj}^{(R)} - \sum_{(v,R)} f_{uv}^{(R)} = b_j \forall j,$$

where b_j denotes the net supply or demand at node j (typically zero). For efficient computation, we transform the above bounded problem into a minimum-cost maximum-flow (MCMF) formulation. Specifically, we define residual variables

$$f_{ij}^{(R)} = f_{ij}^{(R)} - L_{ij}^{(R)}, 0 \leq f_{ij}^{(R)} \leq U_{ij}^{(R)} - L_{ij}^{(R)},$$

and adjust node demands accordingly:

$$b'_j = b_j + \sum_{(v,R)} L_{vj}^{(R)} - \sum_{(u,R)} L_{uj}^{(R)}.$$

Here, nodes with $b'_j < 0$ are connected to a super-source, and nodes with $b'_j > 0$ to a super-sink, via auxiliary edges of corresponding capacity and zero cost. A feasible solution exists only if all auxiliary edges are fully saturated. The total minimum cost equals the fixed contribution from all lower bounds plus the minimum cost of the residual flow. The original solution is finally recovered by

$$f_{ij}^{(R)} = f_{ij}^{(R)} + L_{ij}^{(R)}.$$

After computing the receptor-centric edge flows $f_{ij}^{(R)}$, CellSigFlow further refines the signaling resolution to the level of individual L-R. Specifically, for each receptor R in target cell j , the total incoming flow $f_{ij}^{(R)}$ from source cell i is redistributed among all ligands L expressed in i that can bind R . The redistribution weights are determined by a combination of ligand—receptor affinity and ligand gene expression levels, such that

$$f_{ij}^{(L,R)} = \frac{\omega_{ij}^{(L,R)}}{\sum_{L'} \omega_{ij}^{(L',R)}} f_{ij}^{(R)},$$

where

$$\omega_{ij}^{(L,R)} = \psi(\text{affinity}_{ij}^{(L,R)}) \cdot \text{expr}_i(L),$$

with $\psi(\cdot)$ representing the affinity-based weighting function and $\text{expr}_i(L)$ denoting the expression level of ligand L in cell i . This hierarchical redistribution enables CellSigFlow to derive interpretable single-ligand—single-receptor signaling flows while retaining the global optimization structure of the bounded flow model. This bounded flow formulation naturally introduces a competition mechanism: the capacity of each receptor-centric edge is limited, so if one L-R pair occupies a larger portion of the flow, the available capacity for other potential interactions is reduced. Consequently, interactions with stronger affinity or more favorable spatial positioning will ‘seize’ limited signaling resources,

suppressing weaker competitors. Such competition occurs not only among different ligands targeting the same receptor but also in the allocation of signaling from one source cell to multiple neighboring targets. This mechanism effectively reflects the biological resource constraints and exclusivity inherent in molecular and cellular interactions.

Results:

Benchmarking with PDE-based simulations: Direct validation of CCC inference methods is limited by the scarcity of ligand-receptor co-localization data. To establish controlled benchmarks, inspired by COMMOT's simulations, we built a series of partial differential equation (PDE)-based reaction-diffusion models simulating ligand diffusion, production, degradation, and receptor binding dynamics, thereby capturing essential biophysical processes of intercellular signaling. We designed three representative regimes: Type 1, multiple ligands competing for few receptors; Type 2, multiple receptors competing for few ligands; and Type 3, complex interconnected ligand-receptor networks. By varying the numbers and connectivity of ligands and receptors, we generated synthetic networks spanning these scenarios. CellSigFlow was benchmarked against CellChat and COMMOT using multiple evaluation metrics.

The average Spearman correlations (ρ_s) are summarized below:

Type	CellSigFlow ρ_s	COMMOT ρ_s	CellChat ρ_s
Type 1	0.8177	0.7951	0.7504
Type 2	0.8125	0.8060	0.6848
Type 3	0.8151	0.7855	0.6983

Across all regimes, CellSigFlow achieved the highest agreement with ground-truth communication fields, outperforming both alternatives. These PDE-based simulations demonstrate that CellSigFlow more faithfully captures intercellular signaling as continuous spatial flows.

Mouse Olfactory Bulb Data Analysis:

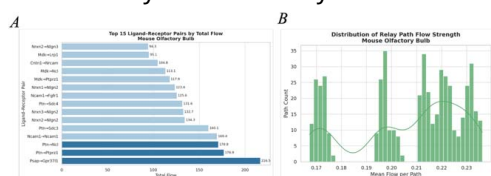


Fig. 1. (A) Top 15 ligand-receptor pairs ranked by total flow. Dark bars highlight dominant signaling consistent with known neuro-glial and growth factor pathways. (B) Distribution of relay path mean flow values. The single peak near 0.2 reflects coherent and consistent two-step signal propagation between neighboring cell clusters.

CellSigFlow was applied to the 10x Genomics Visium spatial transcriptomics data of the mouse olfactory bulb, generating a connected communication network of 191,498 directional edges across 1185 spatial spots. The analysis revealed 591 ligand-receptor pairs and 500 two-step relay paths, outlining coherent signaling hierarchies across the tissue. The Psap-Gpr37l1 axis showed the highest cumulative flow (216.5), consistent

with astrocyte-to-neuron trophic signaling that supports neuronal survival and glial homeostasis. Ptn-Ptprz1, Ptn-Ncl, and Ptn-Sdc3 formed a pleiotrophin (PTN) module promoting neurite outgrowth and synaptic plasticity, while Ncam1-Ncam1 reflected homophilic adhesion crucial for axon guidance and synapse stabilization. Moderate-flow axes such as Fgfr1-Fgfr1 and Sdc4-Fgfr1 indicated fibroblast growth factor (FGF) signaling involved in structural maintenance. Relay path analysis showed consistent mean flow (0.17–0.24), suggesting coordinated rather than random signal propagation. Notably, cascades such as Ptn → Ptprz1 → Cadm3/Nectin1 → Cadm3 and Psap → Gpr37l1 → Nlgn3/Cadm3 linked growth factor and adhesion signaling, mirroring mechanisms of synaptic maturation. Together, these findings demonstrate that CellSigFlow captures both direct and multi-step communication events underlying neuro-glial interactions and synaptic organization in spatial tissue contexts.

Conclusion: In this work, we introduced CellSigFlow, a novel framework that formulates CCC as a bounded minimum-cost flow problem. By explicitly modeling receptor-level constraints, ligand availability, and spatial context, CellSigFlow provides a biologically grounded representation of intercellular signaling. Unlike earlier cluster-level or pairwise methods, CellSigFlow resolves competition among ligands and receptors, enforces realistic upper and lower bounds on signaling flux, and captures multistep relay interactions, thereby offering a richer and more faithful characterization of cellular communication networks.

Through PDE-based simulations, we demonstrated that CellSigFlow outperforms established methods such as CellChat and COMMOT, consistently achieving higher concordance with ground-truth communication fields. This confirms the utility of our flow-based modeling in capturing the dynamic and competitive nature of signaling processes. Application to the mouse olfactory bulb confirmed that CellFlow captures biologically coherent neuron-glia and synaptic signaling, driven by key axes such as Psap-Gpr37l1 and Ptn-Ptprz1. Together, these findings highlight CellSigFlow as a robust and scalable tool for single-cell and spatial transcriptomics analysis. By integrating spatial distance, molecular affinity, and receptor-ligand constraints into a unified optimization problem, CellSigFlow bridges the gap between theoretical modeling and real biological systems. Moving forward, this approach paves the way for systematically dissecting complex signaling landscapes across tissues and diseases, with the potential to uncover novel therapeutic insights and inform subsequent experimental validation.

References

- Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* 2021 Feb 17;12(1):1088.
- Cang Z, Zhao Y, Almet AA, Stabell A, Ramos R, Plikus MV, et al. Screening cell-cell communication in spatial transcriptomics via collective optimal transport. *Nat Methods.* 2023 Feb;20(2):218–28.