ATS Science & Tech

Middle-mile Network Optimization

Minimizing costs while maximizing delivery speed

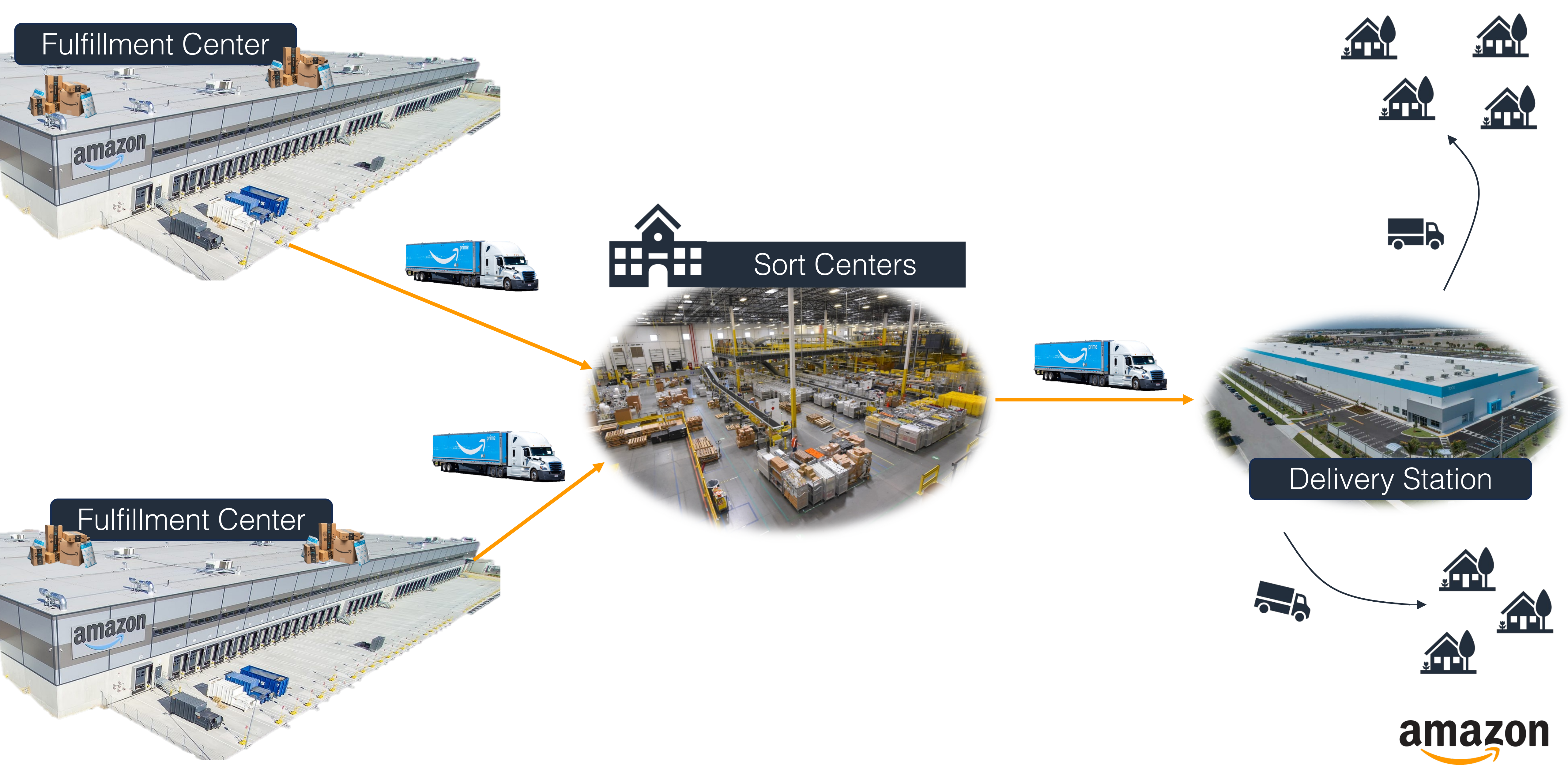
Ugo Rosolia
Research Science Manager

Amazon Transportation Services (ATS), Science & Tech

amazon

Middle-mile Transportation Network

ATS Science & Tech



Middle-mile Optimization

ATS Science & Tech

Key decisions

1. Connectivity, i.e., buildings to connect.
2. Timing, i.e., trucks departure times.

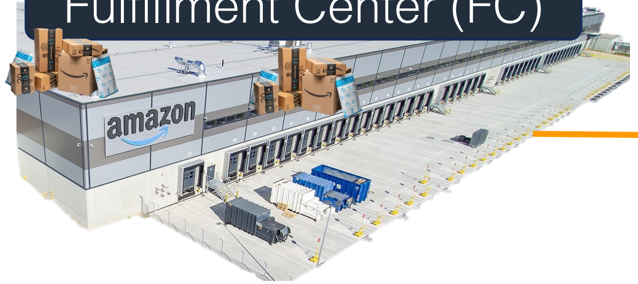
Objectives

1. Reduce cost.
2. Minimize carbon emissions.
3. Maximize delivery speed.

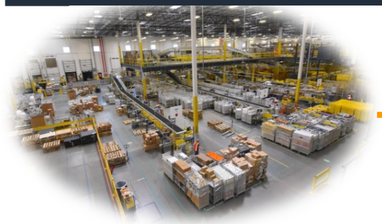
amazon

Middle-mile Optimization

Fulfillment Center (FC)



Sort Centers (SC)



Delivery Station (DS)

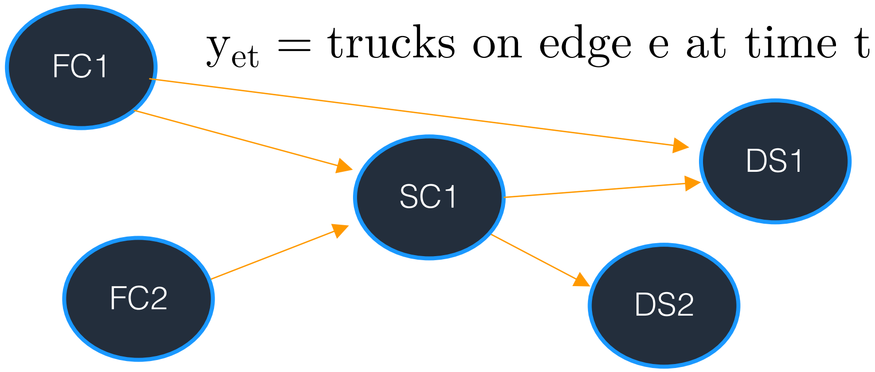
Key decisions

1. Connectivity, i.e., buildings to connect.
2. Timing, i.e., trucks departure times.

Objectives

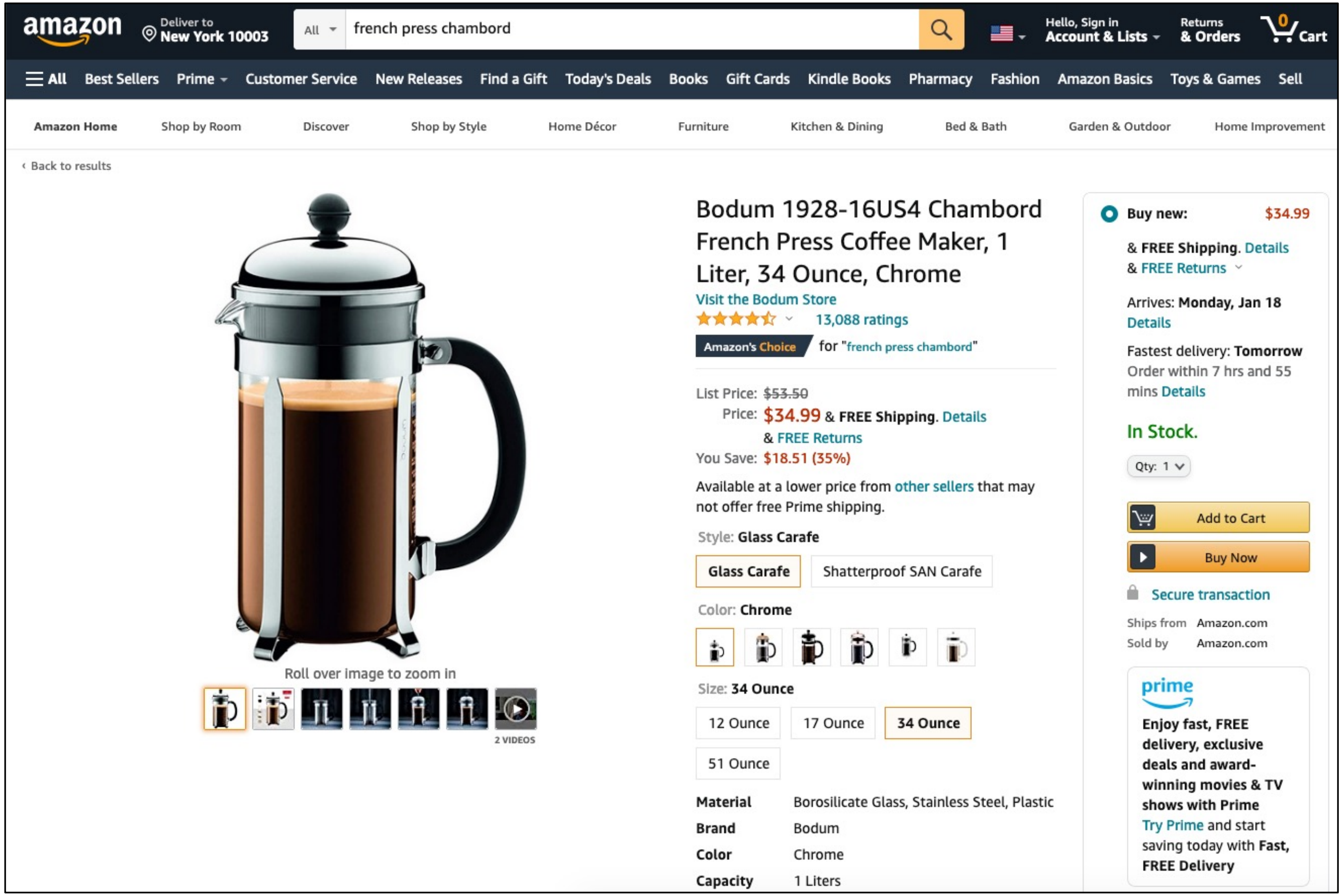
1. Reduce cost.
2. Minimize carbon emissions.
3. Maximize delivery speed.

Time-expanded network flow problem



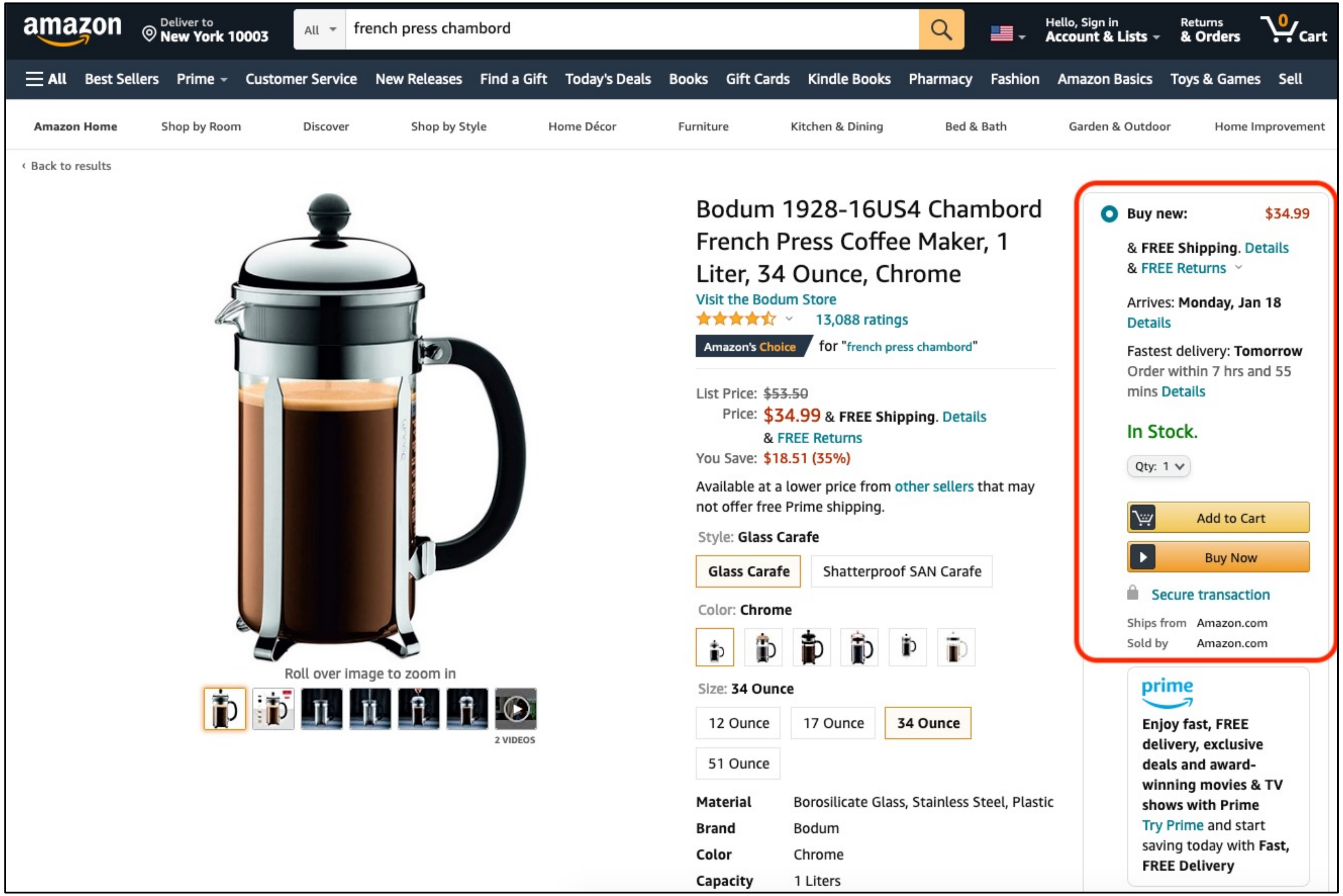
Intractable at Amazon's scale!

Why do we care about speed?



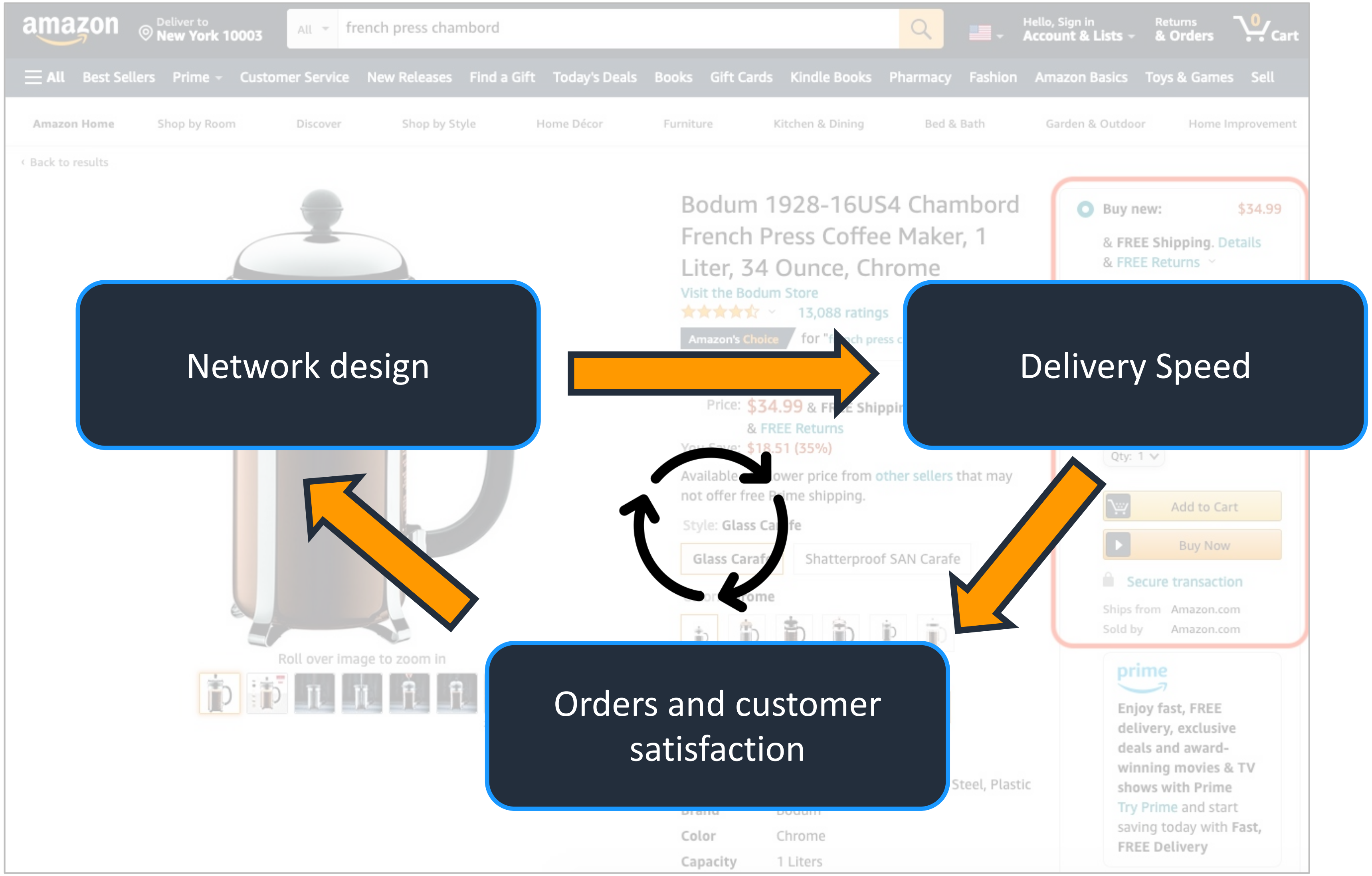
Why do we care about speed?

ATS Science & Tech

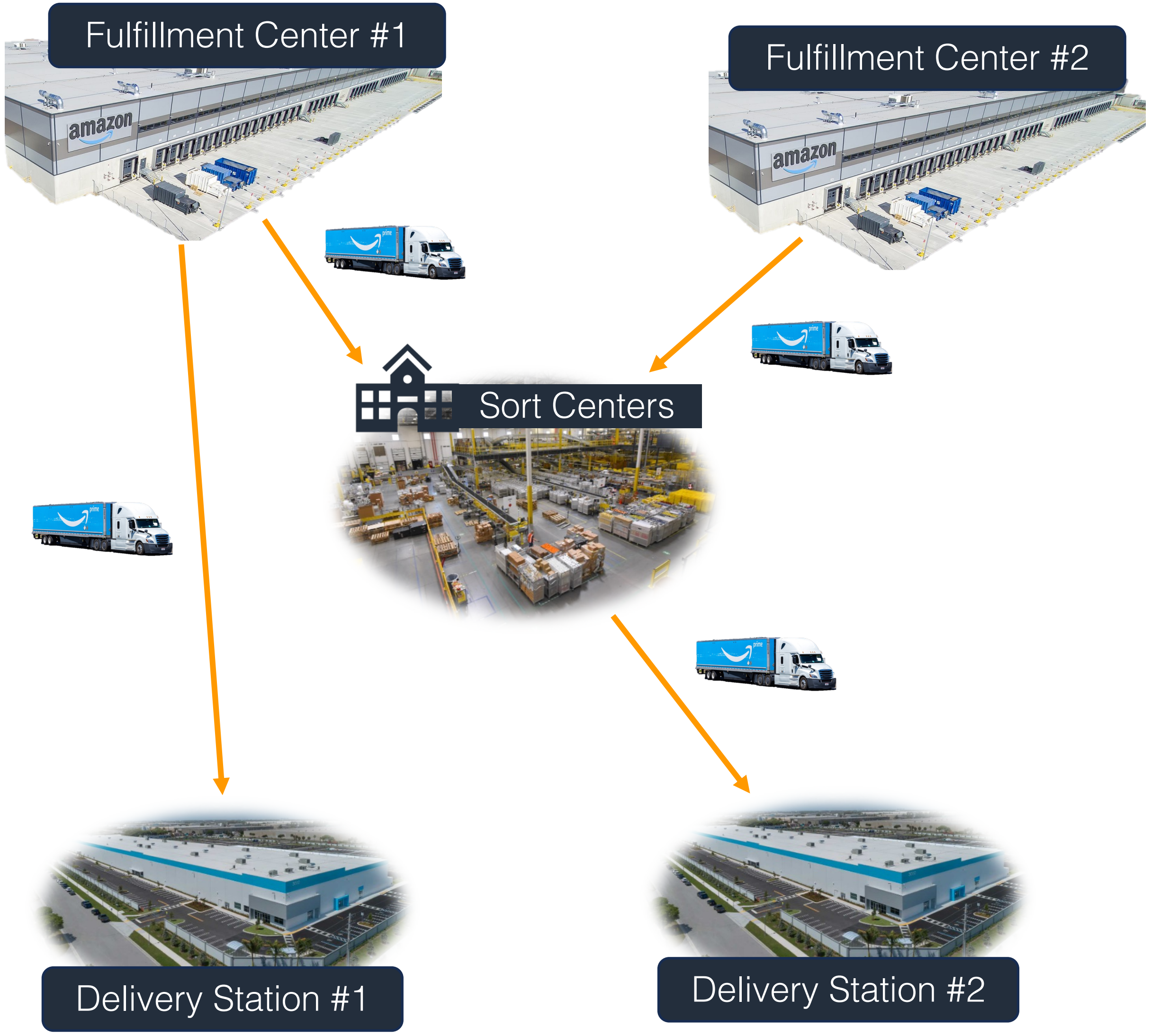


amazon

Why do we care about speed?



Network Design: Connectivity

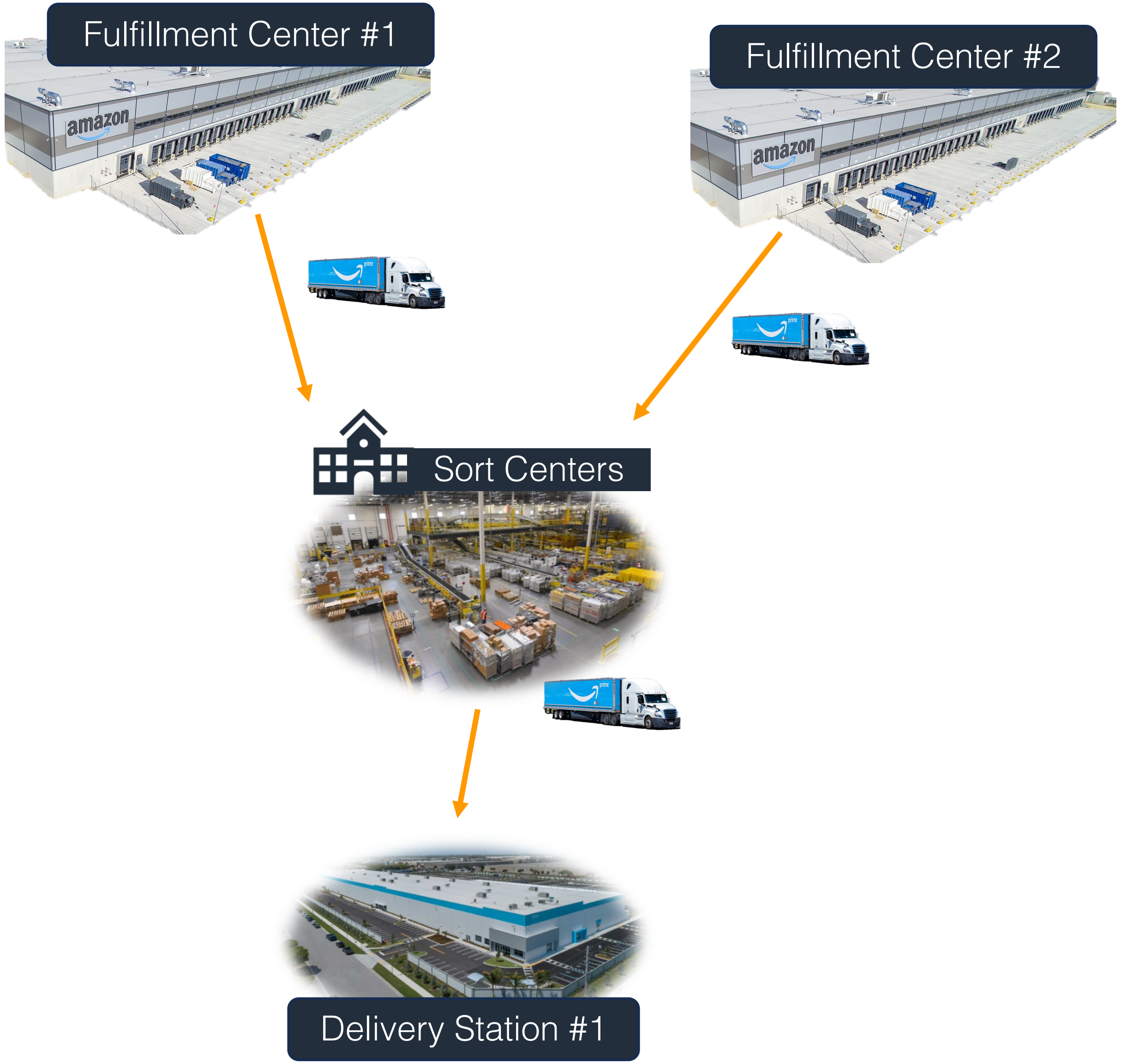


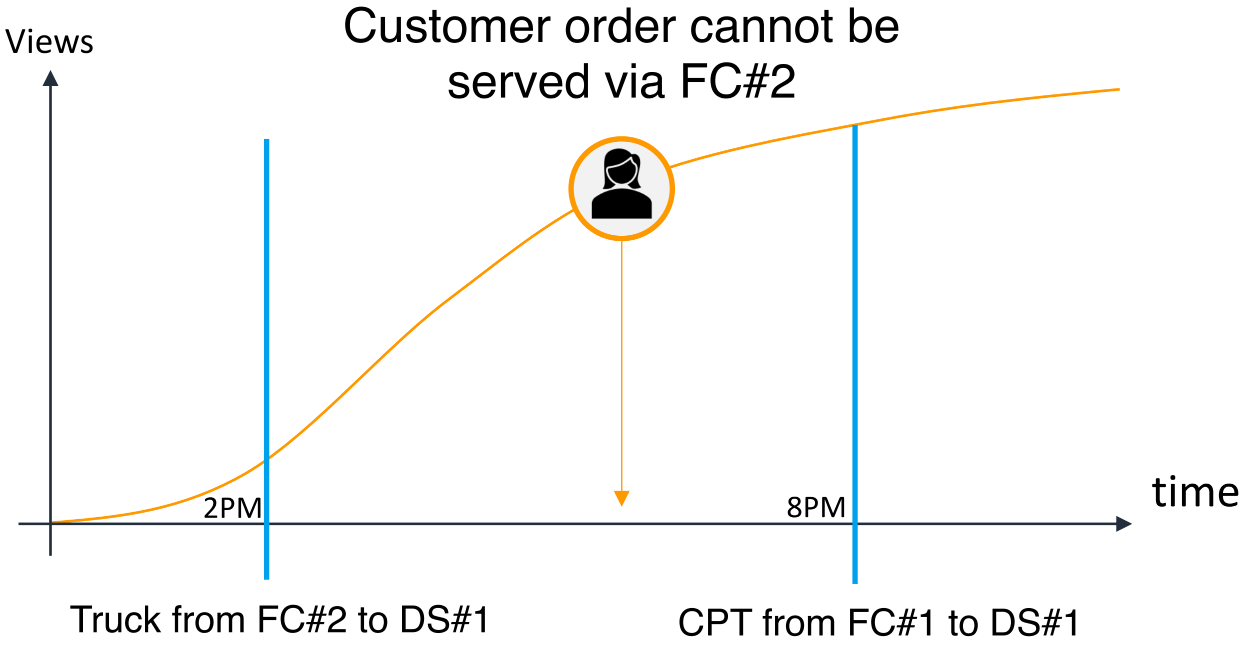
$$\min_{\boldsymbol{p},\boldsymbol{v}} \quad \texttt{NetworkCost}(\boldsymbol{p},\boldsymbol{y})$$
$$\text{s.t.} \quad (\boldsymbol{p},\boldsymbol{y}) \in \texttt{FeasibleNetwork}$$

Path vector variable

Trucks vector variable

Network Design: Timing





$$\max_{\boldsymbol{z}} \quad \texttt{Speed}(\boldsymbol{z})$$

$$\text{s.t.} \quad \boldsymbol{z} \in \texttt{FeasibleSchedule}(\boldsymbol{p})$$

Truck vector variable

Dependency on connectivity

Network Design: The joint problem

$$\begin{aligned} \min_{\boldsymbol{p},\boldsymbol{y},\boldsymbol{z}} \quad & \texttt{NetworkCost}(\boldsymbol{p},\boldsymbol{y}) - \texttt{Speed}(\boldsymbol{z}) \\ \text{s.t.} \quad & (\boldsymbol{p},\boldsymbol{y}) \in \texttt{FeasibleNetwork} \\ & \boldsymbol{z} \in \texttt{FeasibleSchedule}(\boldsymbol{p}) \end{aligned}$$

Why is this problem hard to solve?

(a) Feasibility: We must consider granular non-convex operational constraints, e.g., site opening hours.

(b) Speed objective: Inventory at FCs impacts the speed given by expensive fast connections.

(c) Scale: Billions of variables to model hourly decisions, e.g., when a truck should depart.

(d) Uncertainty: Customers' demand is uncertain, thus we should minimize the expected cost.

Network Design: The joint problem

ATS Science & Tech

$$\begin{aligned} \min_{\boldsymbol{p},\boldsymbol{y},\boldsymbol{z}} \quad & \texttt{NetworkCost}(\boldsymbol{p},\boldsymbol{y}) - \texttt{Speed}(\boldsymbol{z}) \\ \text{s.t.} \quad & (\boldsymbol{p},\boldsymbol{y}) \in \texttt{FeasibleNetwork} \\ & \boldsymbol{z} \in \texttt{FeasibleSchedule}(\boldsymbol{p}) \end{aligned}$$

Why is this problem hard to solve?

(a) Feasibility: We must consider granular non-convex operational constraints, e.g., site opening hours.

(b) Speed objective: Inventory at FCs impacts the speed given by expensive fast connections.

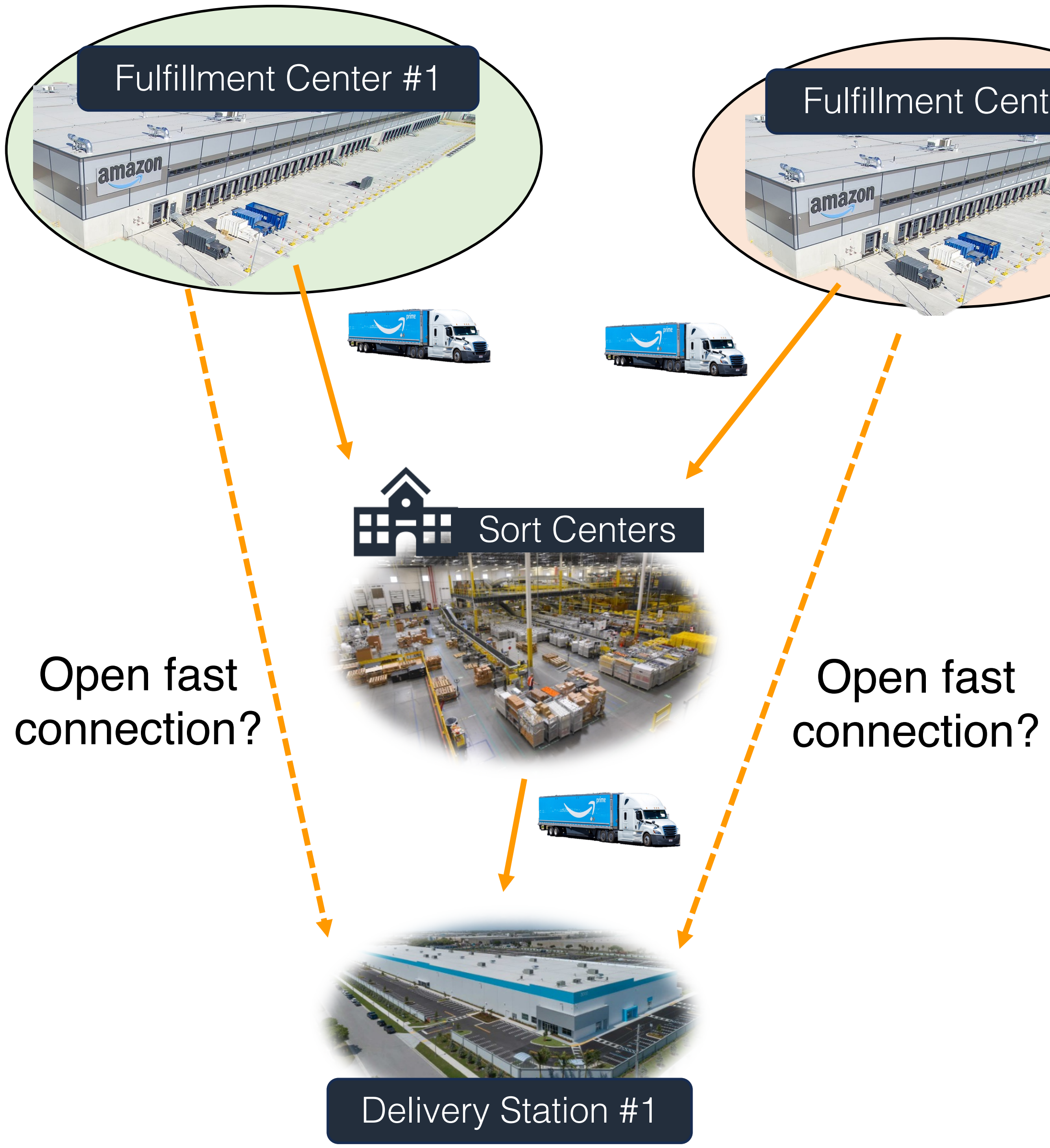
(c) Scale: Billions of variables to model hourly decisions, e.g., when a truck should depart.

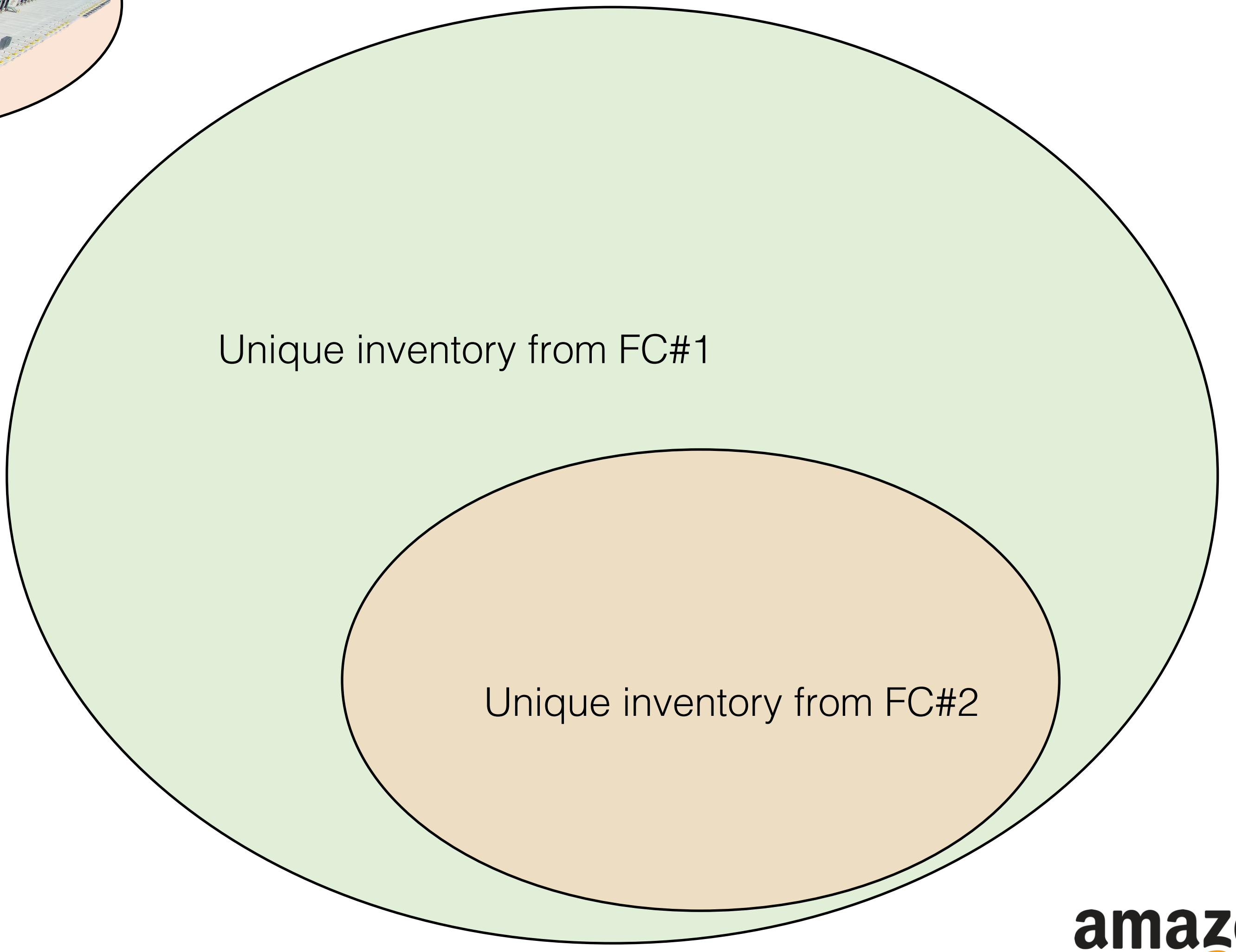
(d) Uncertainty: Customers' demand is uncertain, thus we should minimize the expected cost.

amazon

Speed-aware Network Design

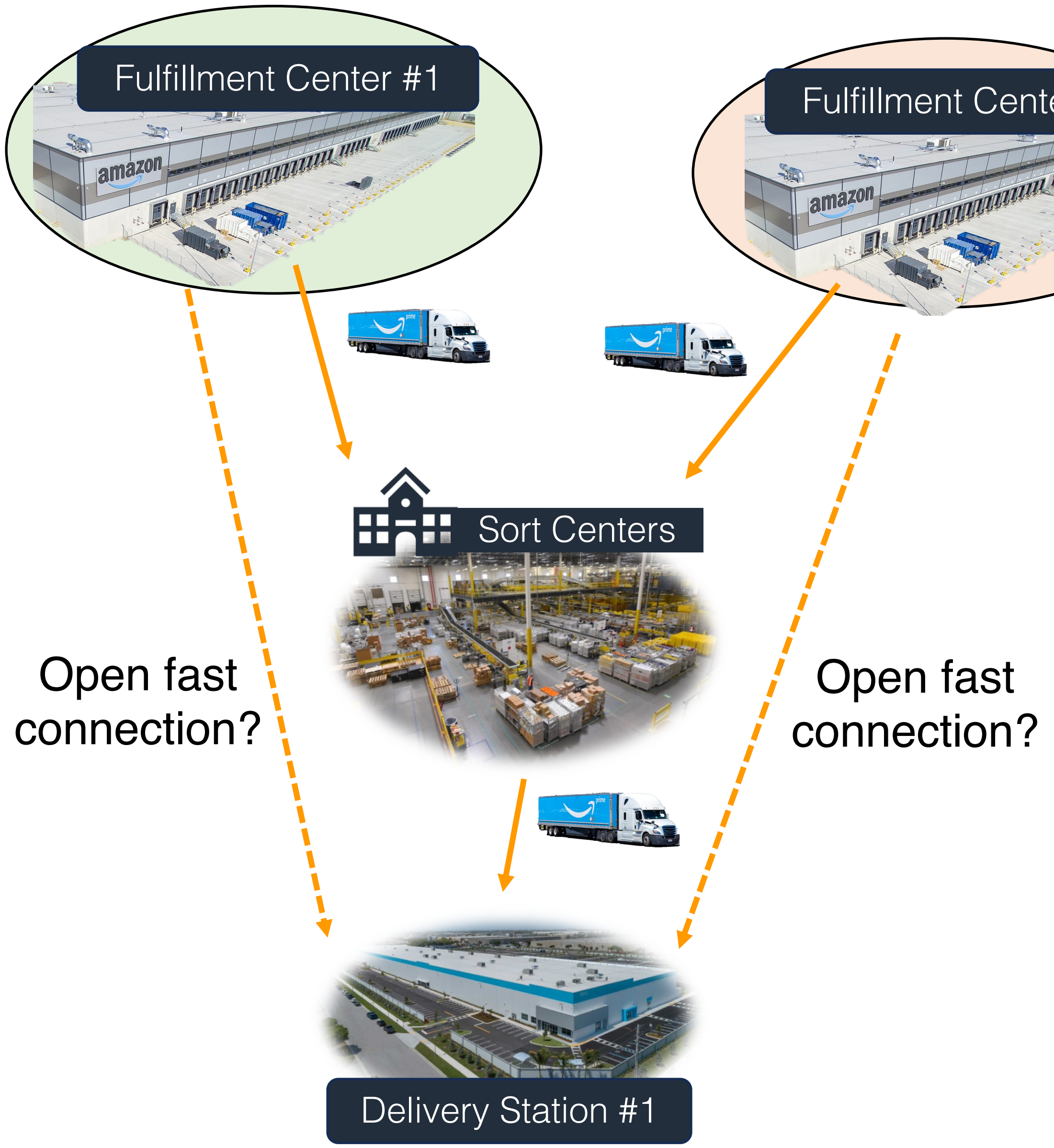
ATS Science & Tech

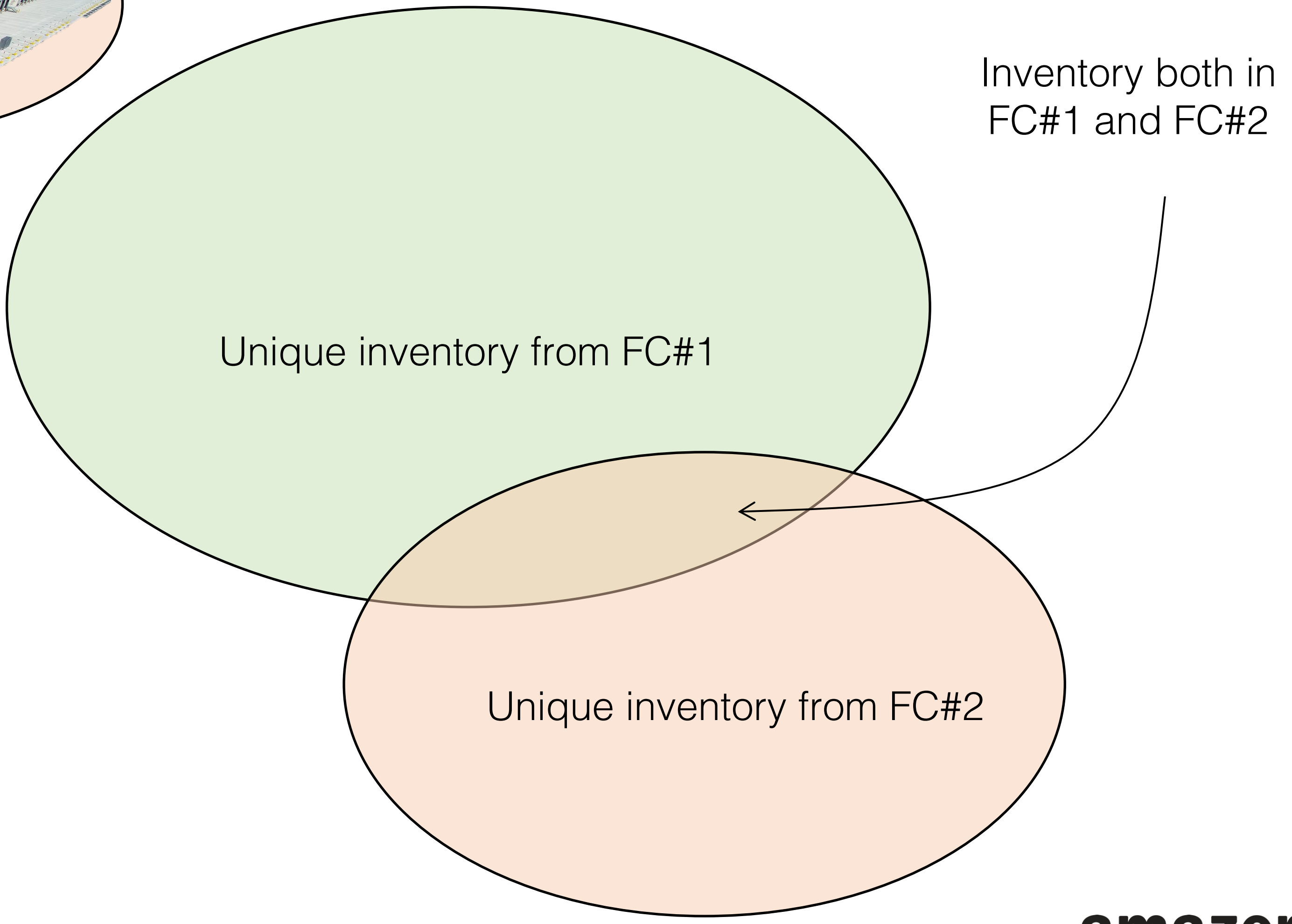




Speed-aware Network Design

ATS Science & Tech

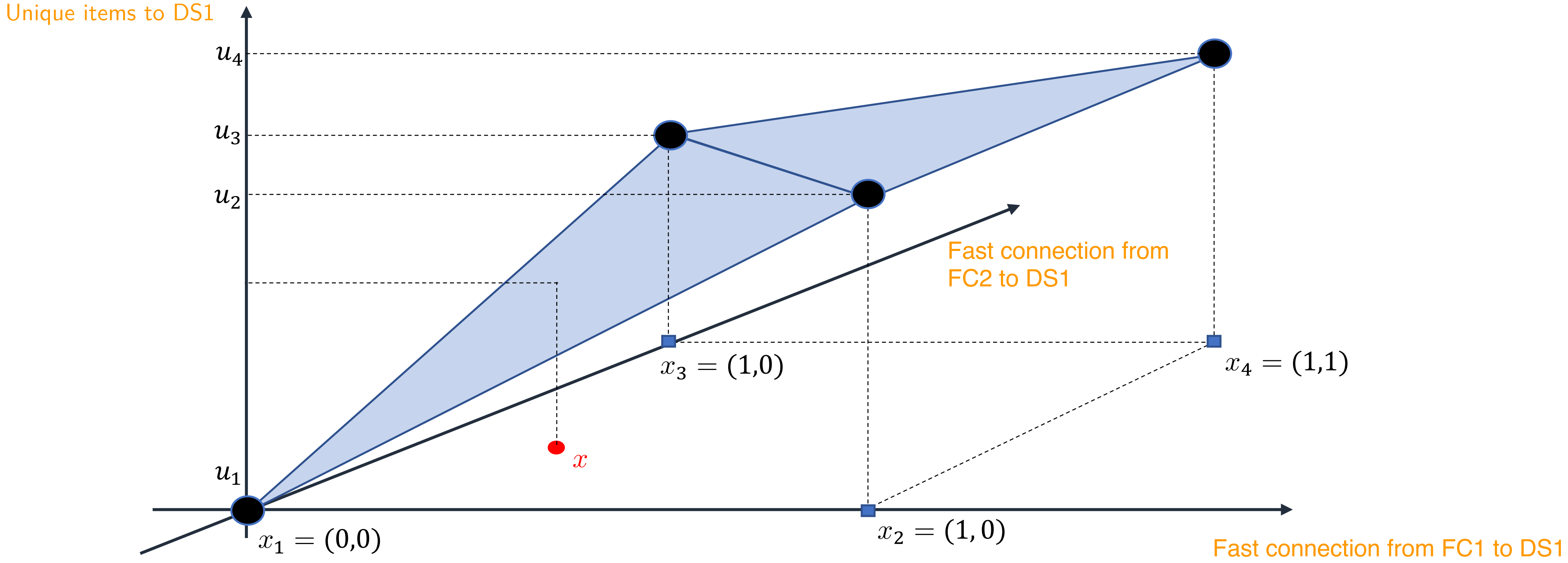




amazon

Speed-aware Network Design

ATS Science & Tech



$$V(x) = \max_{\alpha_i \geq 0} \sum_i \alpha_i u_i$$

$$\text{subject to} \quad \sum_i \alpha_i = 1, \sum_i \alpha_i x_i = x$$

Ugo Rosolia, Marc Bataillou Almagro, George Iosifidis, Amit Kumar, Georgios Paschos, "Speed-aware network design: a parametric optimization approach".

amazon

Network Design: The joint problem

ATS Science & Tech

$$\min_{\boldsymbol{p},\boldsymbol{y},\boldsymbol{z}} \quad \texttt{NetworkCost}(\boldsymbol{p},\boldsymbol{y}) - \texttt{Speed}(\boldsymbol{z})$$
$$\text{s.t.} \quad (\boldsymbol{p},\boldsymbol{y}) \in \texttt{FeasibleNetwork}$$
$$\boldsymbol{z} \in \texttt{FeasibleSchedule}(\boldsymbol{p})$$

Why is this problem hard to solve?

(a) Feasibility: We must consider granular non-convex operational constraints, e.g., site opening hours.

(b) Speed objective: Inventory at FCs impacts the speed given by expensive direct connections.

(c) Scale: Billions of variables to model hourly decisions, e.g., when a truck should depart.

(d) Uncertainty: Customers' demand is uncertain, thus we should minimize the expected cost.

amazon

Solving MIPs with ML + Optimization

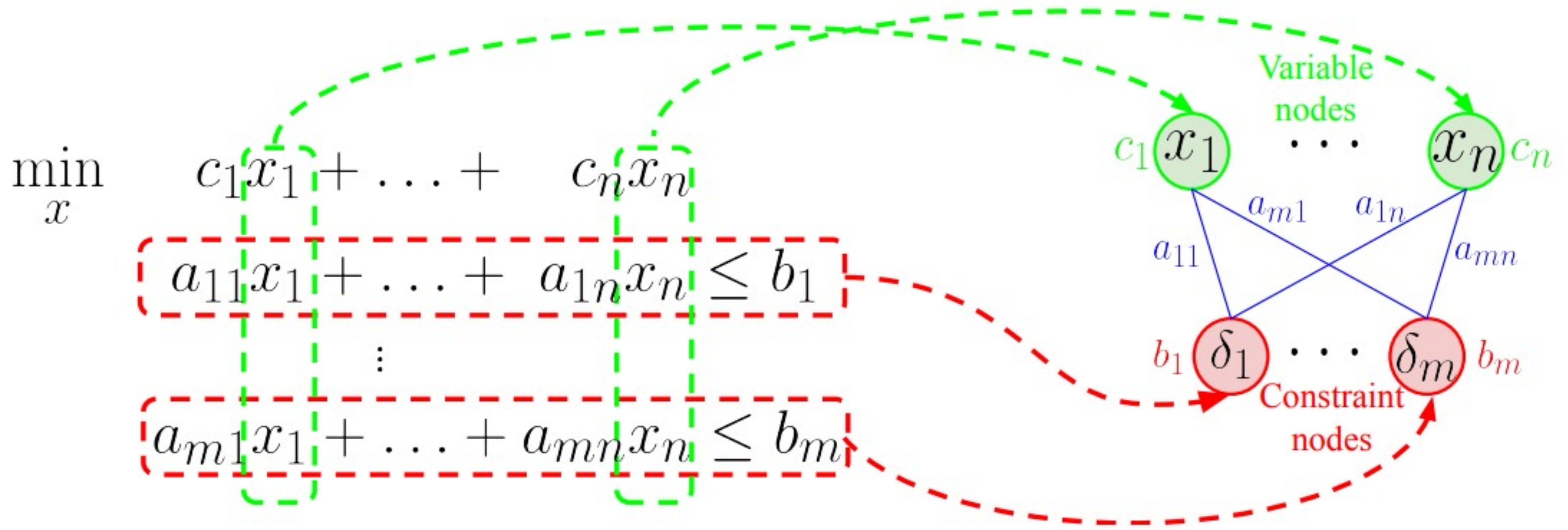
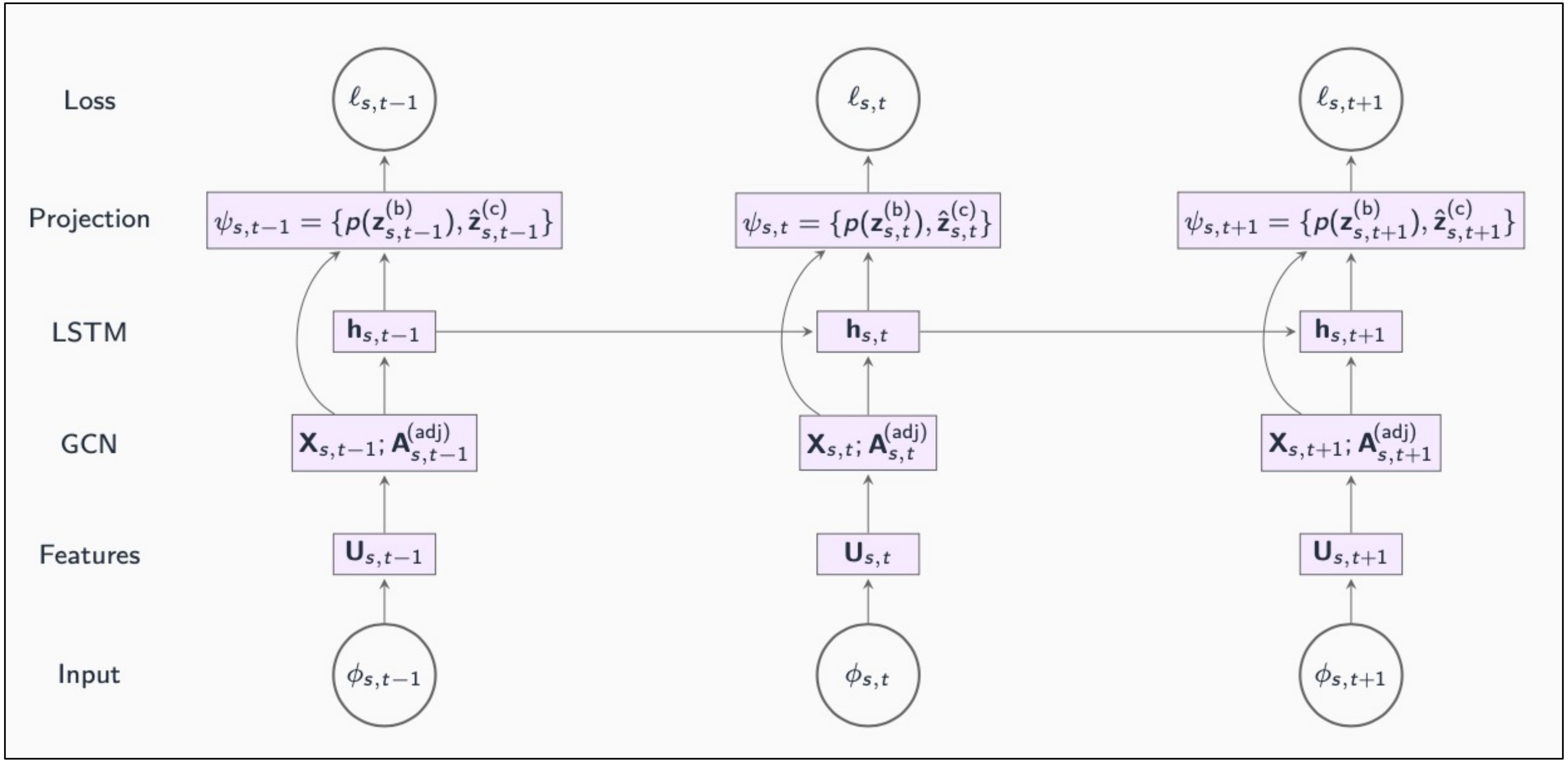


Figure 3 Bipartite graph representation of a MIP used as the input to a neural network. The set of n variables $\{x_1, \ldots, x_n\}$ and the set of m constraints $\{\delta_1, \ldots, \delta_m\}$ form the two sets of nodes of the bipartite graph. The coefficients are encoded as features of the nodes and edges.

Figure from Nair, V., at al (2020). Solving mixed integer programs using neural networks. arXiv preprint arXiv:2012.13349.

Solving MIPs with ML + Optimization

ATS Science & Tech



Key idea:

1. Learn the probability that a variable will be active.
2. Fix variables that are active with high-probability to reduce problem dimensionality

Benidis, K., Rosolia, U., Rangapuram, S., Iosifidis, G., & Paschos, G. (2023). Solving Recurrent MIPs with Semi-supervised Graph Neural Networks. *arXiv preprint arXiv:2302.11992*.

amazon

Solving MIPs with ML + Optimization

ATS Science & Tech

Table 1: Datasets.

	path-routing	facility-location	travelling-salesman	energy-grid	revenue-maximization
# Integer variables	1044	1989	144	1000	10000
# Continuous variables	0	0	0	500	0
# Equality constraints	390	50	36	1	0
# Inequality constraints	36	39	4082	1010	10
Data	real	mixed	synthetic	synthetic	synthetic
Fraction of non-zeros in $\mathbf{A}$	0.0333	0.0226	0.1414	0.6723	0.5011
Fraction of non-zeros in $\mathbf{b}$	1.0	0.5	0.8333	0.9876	1.0
Fraction of non-zeros in $\mathbf{c}$	0.9169	0.9804	0.9996	1.0	1.0
Fraction of non-zeros in $\mathbf{z}^*$	0.4298	0.0278	0.0833	0.7375	0.0166

Benidis, K., Rosolia, U., Rangapuram, S., Iosifidis, G., & Paschos, G. (2023). Solving Recurrent MIPs with Semi-supervised Graph Neural Networks. *arXiv preprint arXiv:2302.11992*.

amazon

Solving MIPs with ML + Optimization

ATS Science & Tech

Table 2: Accuracy percentage (mean ± std) over all instances of MIPnet vs. neural-diving. Bold indicates the best method (higher values are better).

ρ	Method	path-routing	facility-location	travelling-salesman	energy-grid	revenue-maximization
30%	MIPnet	**0.9964±0.0007**	**1.0000±0.0000**	**0.9943±0.0027**	**1.0000±0.0000**	**1.0000±0.0000**
	neural-diving	0.9578±0.0072	**1.0000±0.0000**	0.9642±0.0053	**1.0000±0.0000**	**1.0000±0.0000**
40%	MIPnet	**0.9924±0.0018**	**1.0000±0.0000**	0.9936±0.0024	**1.0000±0.0000**	**0.9997±0.0005**
	neural-diving	0.9083±0.0933	0.9998±0.0002	**0.9961±0.0012**	**1.0000±0.0000**	0.9991±0.0013
50%	MIPnet	**0.9826±0.0022**	**1.0000±0.0000**	**0.9913±0.0029**	**0.9998±0.0000**	0.9997±0.0004
	neural-diving	0.9002±0.0066	0.9991±0.0003	0.9552±0.0046	0.9861±0.0092	**1.0000±0.0000**
60%	MIPnet	**0.9649±0.0044**	**0.9998±0.0001**	**0.9867±0.0041**	**0.9972±0.0010**	**0.9997±0.0005**
	neural-diving	0.9199±0.0271	0.9991±0.0002	0.9761±0.0277	0.9523±0.0030	0.9980±0.0028
70%	MIPnet	**0.9409±0.0048**	**0.9984±0.0005**	**0.9797±0.0046**	**0.9874±0.0011**	0.9995±0.0007
	neural-diving	0.8926±0.0399	0.9964±0.0006	0.9669±0.0220	0.9502±0.0481	**1.0000±0.0000**

Benidis, K., Rosolia, U., Rangapuram, S., Iosifidis, G., & Paschos, G. (2023). Solving Recurrent MIPs with Semi-supervised Graph Neural Networks. *arXiv preprint arXiv:2302.11992*.

amazon

Middle-mile Network Optimization

Minimizing costs while maximizing delivery speed

Ugo Rosolia

Amazon Transportation Services (ATS)
Science & Tech

amazon