

Yang, Xuefei; Ostermeier, Manuel; Hübner, Alexander

**Article — Published Version**

## Winning the race to customers with micro-fulfillment centers: an approach for network planning in quick commerce

Central European Journal of Operations Research

*Suggested Citation:* Yang, Xuefei; Ostermeier, Manuel; Hübner, Alexander (2024) : Winning the race to customers with micro-fulfillment centers: an approach for network planning in quick commerce, Central European Journal of Operations Research, ISSN 1613-9178, Springer Berlin Heidelberg, Berlin/Heidelberg, Vol. 32, Iss. 2, pp. 295-334, <https://doi.org/10.1007/s10100-023-00893-x>

This Version is available at:

<https://hdl.handle.net/10419/314997>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# Winning the race to customers with micro-fulfillment centers: an approach for network planning in quick commerce

Xuefei Yang<sup>1</sup> · Manuel Ostermeier<sup>2</sup> · Alexander Hübner<sup>1</sup> 

Accepted: 7 November 2023 / Published online: 18 January 2024  
© The Author(s) 2024

## Abstract

In 2020, the first quick commerce businesses in grocery retail emerged in the European market. Customers can order online and receive their groceries within 15 min in the best case. The ability to provide short lead times is, therefore, essential. However, the ambitious service promises of quick deliveries further complicate order fulfillment, and many retailers are struggling to achieve profitability. Quick commerce retailers need to establish an efficient network of micro-fulfillment centers (MFCs) in customer proximity, i.e., urban areas, to master these challenges. We address this strategic network problem and formulate it as a location routing problem. This enables us to define the number, location, type, and size of MFCs based on setup, replenishment, order processing, and transportation costs. We solve the problem using a cluster-first-route-second heuristic based on agglomerative clustering to approximate transportation costs. Our numerical experiments show that our heuristic solves the problem effectively and provides efficient decision support for quick commerce retailing. We generate managerial insights by analyzing key aspects of a quick commerce business, such as lead times and problem-specific cost factors. We show, for example, that allowing slightly higher delivery flexibility (e.g., offering extended lead times) enables bundling effects and results in cost savings of 50% or more of fulfillment costs. Furthermore, using multiple small MFCs is more efficient than larger, automated MFCs from a lead time and cost perspective.

**Keywords** Retail · Online grocery · On-demand delivery · Instant delivery · Location routing problem

---

✉ Alexander Hübner  
alexander.huebner@tum.de

<sup>1</sup> Technical University of Munich, Supply and Value Chain Management, Am Essigberg 3, 94315 Straubing, Germany

<sup>2</sup> University of Augsburg, Resilient Operations, Universitätsstraße 12, 86159 Augsburg, Germany

## 1 Introduction

The e-commerce business has been rapidly growing during recent years, and online grocery retailing has been accelerated by the COVID-19 pandemic, ongoing digitization of shopping, and changes in work and lifestyle (see, e.g., Weber 2021). In line with this development, there is also a growing demand for ever faster order fulfillment, such as same-day or even same-hour (Simmons et al. 2022). Orders in this regard comprise products required for instant consumption and have small basket sizes. The particular demand structure and the expected short lead times have facilitated the emergence of new pure-online players such as getir, Flink or Gorillas. They have established a new grocery retail format—known as *quick commerce*—and offer extremely short lead times of quasi-instant deliveries from 15 to 60 min from small inner city depots (Buldeo Rai et al. 2023; Gund and Daniel 2023). This type of business is also referred to as “fast delivery”, “rapid delivery”, “on-demand delivery” (Waßmuth et al. 2023) or “instant delivery” (Buldeo Rai et al. 2022). Today, deliveries within one hour account for 5–10% of Europe’s online grocery market (Sasi 2021). Related business models with promised lead times of a maximum of 30min were valued at around 25 billion \$ in revenue worldwide in 2021, and growth to 72 billion \$ by 2025 is expected (Bommireddipalli 2022). Despite the slower growth and the typical coming and going of new startups in such a dynamic environment (Steinschaden 2022; Partington 2023), the revenue across Europe is valued at around 8 billion \$ in 2021, grew up to 10 billion \$ and is still expected to continue growing to 12 billion \$ by 2025 (Statista 2023b).

This growth is accelerated by relatively limited upfront investments and streamlined processes. Customers usually need first to enter their address when ordering online. The product portfolio is offered if the customer location is within the delivery region. A quick commerce retailer’s product range includes all kinds of categories ranging from fresh, ultra-fresh, chilled, and frozen to ambient products (Ariker 2021) and usually comprises 1500–2000 products (Statista 2022a). However, the average order basket is small and ranges between 5 and 15 items (Statista 2022b), as the orders usually fulfill an immediate need. Instead of standard grocery deliveries with a lead time of 1–3 days, quick commerce does not offer time windows. As the time gap between order placement and receipt is less than 60min, an additional time window specification is almost impossible; instead, a lead time is promised for the delivery (see, e.g., Dablanc et al. 2017; Flink 2023; Getir 2023). After an order is placed, it is instantly transmitted to a depot for order processing. A picker processes the complete order in the depot, and a driver delivers it to the customer. While the picker always processes each order individually (Waßmuth et al. 2023), a driver may wait for additional orders within a given time frame and then start the delivery tour. Lastly, customers must be at home to receive orders (i.e., attended home delivery). The handover of the delivery is the only in-person interaction between the customer and the retailer, motivating the retailer to create a pleasant and timely delivery experience for the customer. Failing to fulfill the service promise is a serious issue due to the core value proposition of fast deliveries.

The main business idea of quick commerce consists of the very short lead time promise combined with many deliveries of small orders. This is a major challenge as last-mile distribution is generally costly (see, e.g., Wollenburg et al. 2018; Waßmuth et al. 2023) and contrasts with the small margins in grocery retail of 2–3% (Klingler et al. 2016). As there is a limited willingness to pay for quick delivery services (Gatta et al. 2021), the price premium is also limited, meaning that all processes must be very cost-efficient. The network and fulfillment processes need to be aligned (Aull et al. 2021) to make the quick commerce shopping experience more attractive for customers than the alternative of shopping in a nearby store without delivery fees. The use of standard shipping services with centralized and large distribution centers (as is common in e-commerce (see, e.g., Hübner et al. 2016, 2019)) is not possible for quick commerce due to the short lead times promise. To enable the ambitious fast delivery, multiple small depots, so-called micro-fulfillment centers (MFCs), must be established in direct customer proximity. Quick commerce retailers rent space in city centers to set up multiple MFCs. From the MFCs, they deliver the orders via bikes or other small vehicles. The small vehicle capacity is a core restriction in quick commerce distribution. This results in networks with many small MFCs and heterogeneous MFC operations. Picking productivity and replenishment operations depend on the available MFC sizes and types, and the lead times and transportation distances to the customers also depend on the MFC locations. The setup of the MFC network is, therefore, pivotal for the long-term success of any quick commerce. However, most quick commerce companies are startups that primarily focus on growth and only attend to efficient operations as a secondary priority (Kale 2021). This often results in a suboptimal fulfillment network. Therefore, it is no surprise that quick commerce retailers may need help to achieve profitable operations, and most businesses suffer from losses or face closure in certain areas (Stone and Davalos 2022). This shows that planning support and efficient solution approaches are needed to enable a sustainable business.

Our work addresses these challenges of quick commerce. It considers a network problem to decide on the number, locations, sizes, and types of MFC, while considering the existing delivery restrictions with short lead times. To establish an efficient distribution network and to assess total fulfillment costs, it is necessary to consider network and routing decisions simultaneously. We formulate the resulting problem as a Capacitated Location Routing Problem with Micro Depots (CLRP\_MD). LRP problems have been widely studied, and there is a plurality of variants (see, e.g., Nagy and Salhi 2007; Prodhon and Prins 2014; Schneider and Drexler 2017). The existing literature does not, however, consider the innovative application in quick commerce when selecting locations, sizes and types of MFCs, and its combination of special delivery requirements and cost structures. Our work addresses these open areas and incorporates the problem characteristic of quick commerce: extremely short lead times without time windows, a decentralized MFC network, differing MFC settings (i.e., storage size and technologies), small order sizes, and very restrictive routing options. We further derive problem-specific costs for quick commerce operations that are considered when determining the MFC network. This includes MFC-specific processing costs (picking and packing), transportation and lateness costs for order fulfillment, and location-dependent replenishment and setup costs. We solve

the resulting problem using a cluster-first-route-second approach tailored to the problem characteristics of quick commerce that approximates transportation costs for a grounded decision on the MFC network structure.

The remainder of this work is structured as follows. Section 2 develops the real-world business problem at hand and the associated network planning problem. Section 3 reviews related literature. Section 4 presents the formal model and the proposed solution approach. Section 5 delineates numerical analyses and derives managerial insights. Last, we summarize our findings in Sect. 6.

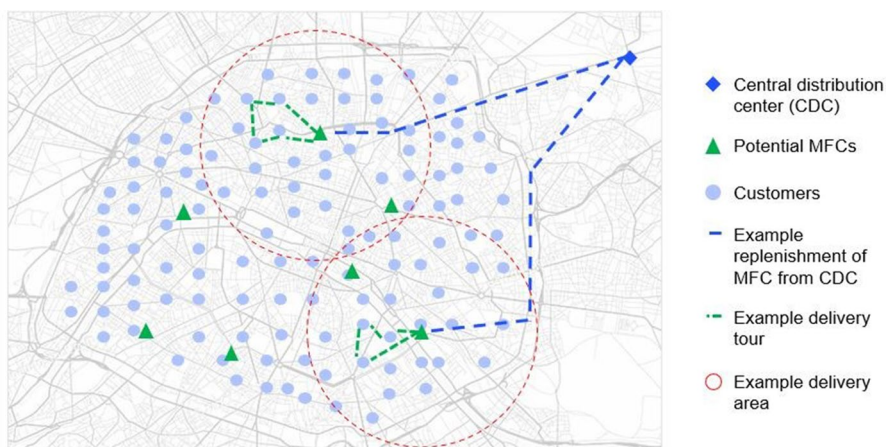
## 2 Description of planning problem

Quick commerce is an innovative concept both in practice and academia. To build a common understanding, this section first derives the fulfillment process and real-world application and then verbalizes the underlying network design problem, its interrelations and related costs.

### 2.1 Description of the real-world business problem

In contrast to the usual mode of traditional retailers with large warehouses outside the city, quick commerce retailers have to set up multiple, spatially distributed MFCs to ensure prompt order fulfillment (Ariker 2021). Only a well-defined network ensures the fast delivery of a variety of products in urban areas. Figure 1 illustrates such a distribution network for quick commerce fulfillment.

Quick commerce retailers operate a two-echelon supply chain. The first echelon is the central distribution center (CDC). The second echelons are the MFCs. In our example, one CDC (diamond) is outside the delivery area. CDCs are organized to efficiently process large volumes (see, e.g., de Koster et al. 2007; Holzapfel et al.



**Fig. 1** Quick commerce fulfillment network illustrating MFC replenishment and customer delivery (example)

2018; Boysen et al. 2019). The inbound process from suppliers is organized via these CDCs, which may serve as classical inventory locations and also partially as cross-docking locations. The CDCs are located outside the urban delivery region due to the space requirements of these warehouses. They are usually unavailable for direct customer delivery and only supply the MFCs. The MFCs (triangles) are set up in inner-city areas such that they are surrounded by the respective delivery areas and customers (dots) and thus in direct customer proximity (Nierynck 2020). The MFCs are frequently replenished from CDCs to ensure high product availability. The MFCs are responsible for order fulfillment and last-mile delivery. Potential locations are limited by the real estate available on the market. Individual MFCs consequently differ significantly, resulting in a heterogeneous network structure. Moreover, MFCs can be set up flexibly in available locations with adapted layouts as no predefined layouts have to be strictly followed compared to classic brick-and-mortar retailers. For example, an MFC can be opened on empty store spaces of 200 to 500 m<sup>2</sup> or even in backrooms of existing local hypermarkets and supermarkets or as attachments of small city stores. When setting up an MFC network, a retailer must decide on the number and locations of the MFCs and the respective available storage size and processing capacity. Regardless of the size, each MFC covers the identical assortment, i.e., the entire product portfolio that is offered in the quick commerce online shop. In smaller-sized MFCs, the replenishment, picking, and packing are exclusively carried out manually. In contrast, in larger MFCs, warehouse automation and robotics to store and retrieve products can be installed to some extent (see, e.g., Boysen et al. 2019). Available technological support ranges from relatively simple conveyor belts purely for material flow in semi-automated goods-to-man systems to automated storage and retrieval systems (see, e.g., Eriksson et al. 2019; Jaghbeer et al. 2020). A simple conveyor belt can be installed with relatively low investment, but productivity benefits are limited. On the other hand, more sophisticated systems can automate more complex processes to improve both productivity and picking capacity but require higher investments (see, e.g., Azadeh et al. 2019). This means that each MFC location needs to be equipped with a certain technology for manual, partially automated, or fully automated order processing, for example. Altogether, the MFC setting at an available location determines the fulfillment (i.e., picking) capacity and the corresponding number of customers that can be supplied.

Quick commerce retailers use defined delivery areas (see red circle in Fig. 1) for each MFC to restrict the delivery radius and to ensure feasible operations (driving range, customer volume). The retailers promise a lead time, which represents the time between the order submission and the receipt by the customer. Due to the small order sizes and tight lead time promises, orders are picked individually quickly without batching across orders (Waßmuth et al. 2023; Buldeo Rai et al. 2023). The fast delivery promise requires that each customer order be picked completely and fulfilled from one MFC. For distribution, picked orders from multiple customers may be bundled within a certain time frame on a tour. Otherwise, when the lead times promised do not allow bundling, the delivery is carried out via direct deliveries. Quick commerce retailers operate their own distribution fleets consisting of small vehicles (e.g., e-bikes, cargo bikes) with limited capacities. This also means the driver returns to the same MFC after completing a tour to take on the next delivery.

Tour sizes in quick commerce are limited due to the short delivery lead times and the limited capacity of the small delivery vehicles. The short lead times restrict the number of customers that can be reached on a single tour due to travel and service times. The vehicle capacity further restricts the number of orders allocated to a tour, even if the lead time allows for bundling.

## 2.2 Decision problem, assumptions and related costs

The resulting network design problem of the quick commerce retailer can be verbalized as follows. The retailer must simultaneously decide on the (i) *number and location* of MFCs and the detailed setting of each MFC. The setting comprises the (ii) *size* and (iii) *type of MFCs*. While the size reflects the available storage space and processing capacity of orders, the type describes the degree of technology and automation that is applied in an MFC, which also affects the processing capacity. The three decisions are interrelated, and resulting tradeoffs must be considered. For example, the decision-making process involves evaluating whether implementing several decentralized, smaller MFCs near customers is more advantageous than deploying fewer, larger, centralized MFCs with the capability for automated picking. This example also shows that the MFC network design needs to include decisions on the (iv) *assignment of customers to the MFCs* for potential order fulfillment and the resulting options for the (v) *vehicle routing* and order delivery. Altogether, the strategic network design problem described can be formally classified as a special variant of a capacitated location routing problem (CLRP) (see, e.g., Nagy and Salhi 2007; Prodhon and Prins 2014; Schneider and Drexler 2017). LRPs combine strategic facility location decisions with operational vehicle routing to supply customers (see, e.g., Daskin 1997). Decisions regarding the placement of facilities and the establishment of efficient transportation routes are intertwined. Research has demonstrated that addressing these aspects separately can lead to excessive overall system costs (see, e.g., Prodhon and Prins 2014). In the standard form, LRPs are formulated as static problems. They are relevant in many applications such as production network planning (see, e.g., Nagy and Salhi 2007; Hasani Goodarzi and Zegordi 2016), disaster relief logistics planning (see, e.g., Oezdamar and Demir 2012; Rath and Gutjahr 2014; Wei et al. 2020) and retail networks (see, e.g., Agatz et al. 2008; Schneider and Drexler 2017). In our problem context, the classical LRP needs to be extended to include decisions on the (ii) MFC size and (iii) MFC type. The network design decision is additionally subject to constraints for picking capacity of MFCs, tight vehicle capacities, and ambitious lead time targets.

The decisions are further based on different cost components that need to be assessed. First of all, each location has different renting and purchasing costs. This is particularly relevant in quick commerce as the MFCs need to be established in costly central urban real estate. Moreover, each location may have different size options, and each MFC can be equipped with different technologies (e.g., semi- or fully-automated picking support). The type selection and, hence, the automation comes along with maintenance and leasing or depreciation costs. Altogether, these costs can be summarized as *fixed setup costs* comprising rental, depreciation, utilities, leasing,



maintenance, and overhead costs that depend on the MFC size, type, and location. Secondly, the MFCs are replenished from the CDC. The size of the MFC determines the required replenishment frequency and volume, its location, and the distance to the CDC that needs to be covered. In addition, the technology installed for inbound replenishment may also impact costs. *Replenishment costs* are consequently location-, size-, and type-specific. Third, there are setup-specific *order processing costs*, as each MFC location, size and type has a different picking productivity. For example, the travel distance of the picker depends on the MFC size and also on the structure of the MFC location (i.e., the MFC type). The latter matters as usually locations in the city center are rented that are not designed for logistics efficiency. The usage of automated picking equipment also influences the productivity of the pickers. The order processing costs account for these picking and packing processes at the MFC and depend on the customer order (e.g., items from different categories) and the MFC location, type, and size. Further, selecting MFC locations also determines the travel distances to customers and the possible routing options. The routing options are limited by tight vehicle capacities and impacted by lead times. The resulting *transportation costs* reflect the travel distances and times that need to be covered within the network (i.e., MFC-customer and customer-customer distances). They account for energy costs, drivers' salaries and vehicle usage. Finally, quick commerce operates with ambitious lead times. This may only be violated in favor of a better assignment solution. Late deliveries are possible to allow for more flexible assignments. Customers expect delivery within the lead time promised but usually accept delays to a certain degree as products are ordered for immediate consumption (Gatta et al. 2021). The lateness leads to future customer churn or financial compensation, which is reflected in *lateness costs*.

Table 1 summarizes the decisions to be made in our special variant of a CLRP and highlights the corresponding tradeoffs by matching the decisions (in columns) with the related costs (in lines). For example, assigning more customers to one MFC allows the setup of a larger MFC and investment in automation technologies. This is beneficial for the order processing costs but increases the fixed setup costs and, in particular, the travel distances and transportation costs to customers.

**Table 1** Summary of network design decisions and impacted costs

Costs   Decisions	(i) MFC location	(ii) MFC size	(iii) MFC type	(iv) Assignment <sup>1</sup>	(v) Routing <sup>2</sup>
Fixed setup costs	✓	✓	✓		
Replenishment costs	✓	✓	✓		
Order processing costs	✓	✓	✓	✓	
Transportation costs	✓			✓	✓
Lateness costs	✓			✓	✓

✓: Direct dependency of costs on decision

<sup>1</sup>Assignment of customers to MFCs

<sup>2</sup>Vehicle routing (i.e., assignment of customers to tours and sequencing of the tour)



### 3 Review of related literature

Due to the recent development of quick commerce retailers, there is only a small amount of academic literature for this novel application. While this small group of contributions establishes quick commerce as an intriguing new retail format (see, e.g., Dablanc et al. 2017; Gatta et al. 2021; Singh and Liébana-Cabanillas 2022), they do not analyze fulfillment models and network design. Our problem is mainly related to network design problems of e-commerce retailers where locations need to be set up to process orders from online channels and deliver the orders to the customers' homes. We analyzed literature reviews in the e-commerce context (see, e.g., Agatz et al. (2008), Melacini et al. (2018), Caro et al. (2020), Hübner et al. (2022), Waßmuth et al. (2023)) to identify related problems. We highlight the contributions in the following and the relationships in a summarizing table at the end of this section.

In the first set of papers, the authors develop models to integrate stores into fulfillment systems. These publications relate to our problem setting, as decentralized locations are applied to complete customer orders. Aksentev and Altinkemer (2008) is the first contribution that investigates enabling physical stores for online order processing. They formulate a static LRP to determine a subset of stores that should be turned into fulfillment depots and solve it with a Lagrangian relaxation. Each customer order adheres to a strict deadline. The model considers setup and transportation costs based on the specific routing of customer orders. However, fulfillment capacities and location-specific order processing costs are not taken into account. Extending this idea to both stores and depots, Bretthauer et al. (2010) choose a subset of existing stores and depots to fulfill online orders. Their static model includes setup and processing costs and bases the transportation costs on direct shipments to customers. The problem is solved with a branch-and-bound for small-scale problems. Ishfaq and Bajwa (2019) evaluate the profitability of fulfillment from omnichannel DCs, e-commerce DCs or stores. They identify under which circumstances DCs to open and how many orders should be processed at the DCs and stores. The dynamic choice of fulfillment location is made through an outer approximation of lower and upper bounds of the objective function. Fixed operating costs of DCs, order processing costs and direct shipment costs to customers are included in their model. Also exploring the opportunity to use stores and DCs for order fulfillment, Arslan et al. (2021) dynamically assign customer demands to a given set of DCs and stores to minimize costs while respecting a delivery deadline. Time constraints consider processing and fulfillment times. If a delivery cannot be fulfilled within the given deadline, this is penalized as lost sales. Replenishment, setup and order processing costs are included in the model. Transportation costs are calculated as direct delivery from the depot to customers. Amongst existing stores and DCs, Dethlefs et al. (2022) assign available orders to fulfillment locations. They apply a cluster-first-route-second heuristic for the assignment and routing problem. The authors consider differing order processing costs in various fulfillment locations and the transportation costs based on routing with a maximum route duration constraint as a

reflection of rapid delivery. Extending the idea of using stores for order pickup by customers, Janjevic et al. (2021) optimize an omnichannel distribution network by changing the number of DCs and order pickup points for customers. This static LRP represents a three-tiered multi-modal network. While capacities and costs of depots are part of the model, the transportation costs are estimated with a route cost formula based on a continuous approximation. The time aspect is included as a maximum duration of each tour without a lead time restriction. Mahar et al. (2022) apply a dynamic program to study a similar problem and develop a dynamic order allocation policy for fulfilling online orders from stores. They consider average processing costs per order, direct shipment costs by carrier zone from store to customer, and focus on expected inventory costs. Delivery time aspects are not considered.

The second set of papers analyzes establishing new central or regional depots for online fulfillment. For example, Pulido et al. (2015) develop a logistics cost function that includes setup, processing, inventory and transportation costs. They determine the optimal number of warehouses for the customer density, the routing zones for each demand period, and the number of orders to be consolidated into a route. To estimate the transportation costs, they adopt the continuous approximation model of Daganzo (1984) for retailers facing short time windows with different urgency. The authors consider up to two-hour deadlines by limiting the number of customers per tour. Actual travel and service times are not taken into account. Acimovic and Graves (2015) introduce a multi-product model that dynamically chooses the best fulfillment location for each product amongst a set of DCs. DC-specific fixed and variable costs and lead time are not integrated in the approximate dynamic program. Rahmani et al. (2016) consider a two-echelon delivery network of a retailer. The first echelon is a CDC that supplies the second echelon of processing depots. They use a clustering-based solution approach to solve the static LRP. Setup costs of a depot and transportation costs based on actual routing are included, while order processing costs are neglected. The vehicles are capacitated, but depots are unconstrained, and delivery deadlines are expressed through time limits for route duration. Millstein and Campbell (2018) present a case study where a retailer chooses the optimal locations for depots to fulfill online orders and to replenish stores. The static location and allocation problem is implemented as a mixed-integer problem (MIP) in a solver with setup costs for various depot capacities. Also working with direct shipments, Kang et al. (2022) improve the delivery stations network for a retailer in cities. Delivery stations are small cross-docking hubs before the last-mile delivery to customers. These delivery stations are selected based on optimizing fixed setup and delivery costs. A neighborhood heuristic is applied to solve the dynamic problem. Millstein et al. (2022) formulate a MIP to determine depot location and capacities to fulfill online and offline orders. A solver is applied to the MIP. They account for fixed costs and picking costs. Finally, Vazquez-Noguerol et al. (2022) introduce a static multi-objective location-allocation model that optimizes the overall operational efficiency while minimizing picking and delivery costs and balancing workload when allocating orders to depots. Small-scale problems are solved using a MIP solver. The authors consider time windows. All papers in the second stream model the transportation costs as direct shipments to customers. They do not build

**Table 2** Related literature on e-commerce fulfillment networks

Paper (chronological)	Decisions		Costs					Constraints			Method <sup>10</sup>			
	Loc. <sup>1</sup>	Type <sup>2</sup>	Size <sup>3</sup>	Assg. <sup>4</sup>	Rout. <sup>5</sup>	Set.	Rep.	Proc. <sup>6</sup>	Trans. <sup>7</sup>	Late <sup>8</sup>		Dep.	Veh.	Time <sup>9</sup>
Aksen and Altinkemer (2008)	✓		(✓)	✓	✓	✓	✓		✓			✓	✓	LR
Brethauer et al. (2010)	✓			✓		✓		✓	*		✓			B & B
Acimovic and Graves (2015)	✓			✓					*					ADP
Pulido et al. (2015)	(✓)			✓	(✓)	✓			(✓)				(✓)	CA
Rahmani et al. (2016)	✓			✓	✓	✓	✓		✓			✓	(✓)	NN, CFRS
Millstein and Campbell (2018)	✓		✓	✓		✓			*		✓			MIP
Ishfaq and Bajwa (2019)	✓			✓		✓	✓	✓	*					OA
Arslan et al. (2021)	✓			✓		✓	✓	✓	*	✓			✓	LSM
Janjevic et al. (2021)	✓			✓	(✓)	✓	✓	✓	(✓)		✓		(✓)	CA
Mahar et al. (2022)	✓			✓		✓		✓	*		✓			DP
Dethlefs et al. (2022)	✓			✓	✓			✓	✓		✓		(✓)	CFRS
Kang et al. (2022)	✓			✓		✓			*		✓			NR
Millstein et al. (2022)	✓		✓	✓		✓	✓	✓	*		✓			MIP
Vazquez-Noguerol et al. (2022)	✓			✓		✓		✓	*		✓		✓	MIP
This paper	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	(✓)	CFRS

<sup>1</sup>✓ Number of required depots and their locations are determined; (✓) Location only approximated<sup>2</sup>✓ Determination of depot type (e.g., automated)<sup>3</sup>✓ Selection of capacity; (✓) derivation of depot capacity by customer assignment<sup>4</sup>✓ Customer assignment to depot<sup>5</sup>✓ Assignment of customers to tours and sequencing of customers; (✓) only approximation of tours (e.g., without sequencing)<sup>6</sup>✓ Variable order processing costs at depot (e.g., picking and packing)<sup>7</sup>✓ Transportation costs to customers are calculated based on the solution of a TSP; (✓) transportation costs to customers are approximated without actual routing (e.g., using continuous approximation); \* only considering direct shipments to customers without routing<sup>8</sup>✓ Penalty costs for late delivery<sup>9</sup>✓ Limits for customer lead time for individual orders are part of the problem setting; (✓) General time limits such as total route length or shift length are part of the problem setting<sup>10</sup>Solution approach: (ADP) (Approximate) dynamic programming; B&B Branch and bound; CA Continuous approximation; CFRS Cluster-first-route-second; MIP Mixed integer program solved with solver; LR Lagrangian relaxation; LSM L-shaped method; NV Nearest neighbour; NR Neighborhood restrictions; OA Outer approximation

vehicle routes for the joint delivery of multiple customers. The only exception is Pulido et al. (2015) that apply a route cost estimation with a variant of the continuous approximation of Daganzo (1984).

*Summary and contribution.* Table 2 summarizes the related literature determining locations and order allocations in an e-commerce context. It highlights the different decision scope of the papers. The common denominator of the static and dynamic models is selecting locations and allocating orders to locations. Therefore, all these papers constitute different variants of a LRP. However, the majority does not include the size and type selection. In many cases, transportation is simplified by assuming direct shipment costs without routing customers on one tour. Furthermore, it becomes evident that decision-relevant costs for quick commerce fulfillment are only partially considered. In particular, the considerations of actual costs for the routing (instead of only approximating it with direct deliveries) and location-, type-, and size-specific setup, replenishment and processing costs, and penalty costs for late deliveries are not comprehensively integrated. Furthermore, the tight capacity restrictions for vehicles and at the depot are only partially considered. Regarding the solution approach, no established exact or heuristics approach is widely used in e-commerce network problems. To fill this gap, our paper aims to consider capacities, costs, routing and delivery time aspirations. Our model includes location-specific setup, replenishment, order processing, transportation and lateness costs.

## 4 Model and solution approach

This section first introduces the mathematical formulation of the CLRP\_MD (Sect. 4.1), and then presents our CFRS specialized heuristic (Sect. 4.2).

### 4.1 The capacitated location routing problem with micro-depots

The CLRP\_MD minimizes the total Fulfillment Costs (FC) by determining the number, locations, sizes, and types of MFCs, allocating customers to MFCs, assigning customers to delivery tours, and determining the sequence of each tour. Table 3 summarizes the notation used.

The CLRP\_MD is formulated on a complete undirected, weighted graph  $G = (N, E)$ , where  $N = \{1, \dots, \bar{l}, \bar{l} + 1, \dots, n\}$  represents all  $n$  locations of the network, and  $E = \{\{i, j\} : i, j \in N, j > i\}$  the set of edges connecting all nodes. Each edge is associated with a non-negative travel time  $d_{ij}$  to account for the driving time between locations  $i, j \in N$ . Furthermore, the delivery at customer  $i$  is associated with a service time  $s_i$  accounting for customer interaction (order receipt, payment) in attended home delivery. The set of locations  $N$  comprises the set of potential MFC locations  $L = \{1, \dots, \bar{l}\}$ , with  $\bar{l} < n - 1$ , and the set of customer locations  $C = \{\bar{l} + 1, \dots, n\}$ , with  $N = L \cup C$ . The subset  $L_i$  represents the potential MFC locations a customer  $i \in C$  can be assigned to.  $L_i$  reflects the requirement that MFCs can only fulfill customer orders within a predetermined maximum

**Table 3** Notation of the CLRP\_MD

<i>Sets</i>	
$C$	Customer locations, with $i \in C$
$L$	Potential locations for MFCs, with $l \in L$
$L_i$	Subset of potential MFC locations where a customer $i$ can be assigned to, with $L_i \subseteq L$
$M$	Sizes of MFCs, with $m \in M$
$N$	All locations, with $i, j \in N$ and $N = L \cup C$
$T$	Types of MFCs, with $t \in T$
$V$	Homogeneous vehicles, with $v \in V$
<i>Parameters</i>	
$c_{lmt}^{\text{fix}}$	Fixed setup costs of MFC size $m$ and type $t$ at location $l$ , $m \in M, t \in T, l \in L$
$c_{lmt}^{\text{rep}}$	Inbound replenishment costs of MFC size $m$ and type $t$ at location $l$ , $m \in M, t \in T, l \in L$
$c_{ilmt}^{\text{proc}}$	Processing costs of the order by customer $i$ at MFC location $l$ with size $m$ and type $t$ , $i \in C, m \in M, t \in T, l \in L$
$c_{ij}^{\text{trans}}$	Transportation costs between locations $i, j$ , $i, j \in N$
$c_i^{\text{late}}$	Penalty costs for late delivery at customer $i$ , $i \in C$
$d_{ij}$	Travel time from location $i$ to $j$ , $i, j \in N$ (in time units)
$D$	Lead time for customer deliveries (in time units)
$p_i$	Picking volume of the order by customer $i$ , $i \in C$ (in picking units)
$q_i$	Transportation volume of the order by customer $i$ , $i \in C$ (in transportation units)
$Q$	Maximum transportation capacity of a vehicle (in transportation units)
$s_i$	Service duration at customer location $i$ , $i \in C$ (in time units)
$S_{mt}$	Maximum picking capacity of MFC size $m$ and type $t$ , $m \in M, t \in T$ (in picking units)
<i>Variables</i>	
$r_i$	Continuous variable, indicating lateness (delay) of the delivery to customer $i$ , $i \in C$
$w_{iv}$	Continuous variable, indicating the arrival time of vehicle $v$ at location $i$ , $i \in N, v \in V$
$x_{ijv}$	Binary variable, indicating whether the vehicle $v$ directly travels from $i$ to $j$ , $i, j \in N, v \in V$
$y_{lmt}$	Binary variable, indicating whether location $l$ will be chosen to build an MFC of size $m$ and type $t$ , $m \in M, t \in T, l \in L$
$z_{ilmt}$	Binary variable, indicating whether customer $i$ is assigned to MFC size $m$ and type $t$ at location $l$ , $i \in C, m \in M, t \in T, l \in L$

distance (see also Fig. 1). Quick commerce customers desire order receipt as soon as possible. This is respected with a promised lead time  $D$  for customer deliveries. Deliveries aim to arrive at customers at this upper time limit for the travel and service time; otherwise, a delivery is considered a late delivery. Set  $M$  represents the potential sizes of MFCs and set  $T$  the potential types of MFCs at each location  $l \in L$ . The MFCs have a limited picking capacity  $S_{mt}$  that depends on their size  $m$  and type  $t$ . Further, a sufficiently large set of homogeneous vehicles  $v \in V$  with capacity  $Q$  is available for the deliveries. Each customer  $i \in C$  in the network considered submits a single order. Each customer's order is associated with a picking volume  $p_i$  at the MFC and transportation volume  $q_i$  for the delivery.

All MFCs are replenished from a CDC with full truckloads. This imposes inbound replenishment costs  $c_{lmt}^{\text{rep}}$  of each MFC that depend on the location (i.e., distance to the CDC), the size, and the type of the MFC. The MFC type and size determine the required replenishment frequency and the inbound handling processes (e.g., processing capabilities and automation). Setting up an MFC imposes the location-, size- and type-dependent fixed setup costs that are denoted by  $c_{lmt}^{\text{fix}}$ . The processing of the order by customer  $i$  and respective picking and packing costs in an MFC are denoted by  $c_{ilmt}^{\text{proc}}$ . Picking and packing costs depend on the customer's order (e.g., number of products, mix of customer orders) and on the MFC setting chosen (location, size and type). The transportation costs are denoted by  $c_{ij}^{\text{trans}}$  as the costs for travelling between locations  $i$  and  $j$ . Finally, we apply lateness costs  $c_i^{\text{late}}$  if the aspired lead time  $D$  is not met. Please note that the decision-relevant costs are derived in the analysis of Sect. 2.

The following decision variables represent the decisions to be made. The binary variable  $y_{lmt}$  indicates whether location  $l$  will be used to open an MFC of size  $m$  and type  $t$ . The binary variable  $z_{ilmt}$  indicates whether customer  $i$  is assigned to the MFC at location  $l$ , size  $m$  and type  $t$  ( $z_{ilmt} = 1$ ) or not ( $z_{ilmt} = 0$ ). The binary variable  $x_{ijv}$  indicates whether vehicle  $v$  drives directly from location  $i$  to  $j$ . Two auxiliary variables are applied. The continuous variable  $w_{iv}$  defines the arrival of vehicle  $v$  at customer  $i$ , and the auxiliary variable  $r_i$  indicates the amount of time the arrival time at customer  $i$  exceeds the lead time  $D$ . The CLRP\_MD is formulated as follows.

$$\begin{aligned} \text{Min } FC(x, y, z) = & \sum_{l \in L} \sum_{m \in M} \sum_{t \in T} (c_{lmt}^{\text{fix}} + c_{lmt}^{\text{rep}}) \cdot y_{lmt} + \sum_{i \in C} \sum_{l \in L} \sum_{m \in M} \sum_{t \in T} c_{ilmt}^{\text{proc}} \cdot z_{ilmt} \\ & + \sum_{i \in N} \sum_{j \in N} \sum_{v \in V} c_{ij}^{\text{trans}} \cdot x_{ijv} + \sum_{i \in C} c_i^{\text{late}} \cdot r_i \end{aligned} \quad (1)$$

subject to

$$\sum_{l \in L} \sum_{m \in M} \sum_{t \in T} z_{ilmt} = 1 \quad \forall i \in C \quad (2)$$

$$\sum_{j \in N} \sum_{v \in V} x_{ijv} = 1 \quad \forall i \in C \quad (3)$$

$$\sum_{m \in M} \sum_{t \in T} y_{lmt} \leq 1 \quad \forall l \in L \quad (4)$$

$$\sum_{i \in L} \sum_{j \in C} x_{ijv} \leq 1 \quad \forall v \in V \quad (5)$$

$$\sum_{j \in N} x_{ijv} = \sum_{j \in N} x_{jiv} \quad \forall i \in N, v \in V \quad (6)$$

$$\sum_{h \in N} x_{ihv} + \sum_{h \in N} x_{lhv} \leq 1 + \sum_{m \in M} \sum_{t \in T} z_{ilmt} \quad \forall i \in C, l \in L, v \in V \quad (7)$$

$$w_{lv} = 0 \quad \forall l \in L, v \in V \quad (8)$$

$$w_{iv} + s_i + d_{ij} - 2D \cdot (1 - x_{ijv}) \leq w_{jv} \quad \forall i \in N, j \in C, v \in V \quad (9)$$

$$w_{iv} \leq D + r_i \quad \forall i \in C, v \in V \quad (10)$$

$$\sum_{i \in C} p_i \cdot z_{ilmt} \leq S_{mt} \cdot y_{lmt} \quad \forall l \in L, m \in M, t \in T \quad (11)$$

$$\sum_{i \in C} \sum_{j \in N} q_i \cdot x_{ijv} \leq Q \quad \forall v \in V \quad (12)$$

$$r_i \geq 0 \quad \forall i \in C \quad (13)$$

$$w_{iv} \geq 0 \quad \forall i \in N, v \in V \quad (14)$$

$$x_{ijv} \in \{0, 1\} \quad \forall i, j \in N, v \in V \quad (15)$$

$$y_{lmt} \in \{0, 1\} \quad \forall l \in L, m \in M \quad (16)$$

$$z_{ilmt} \in \{0, 1\} \quad \forall i \in C, l \in L, m \in M \quad (17)$$

The Objective Function (1) minimizes the total fulfillment costs. The first term sums up fixed setup and inbound replenishment costs that depend on the location, size, and type of MFCs selected. The second term calculates the total order processing costs across all MFCs. The third term considers transportation costs for the delivery of customer orders, and the final term calculates the lateness costs. Constraints (2) and (3) ensure that each customer  $i$  is assigned to exactly one MFC and delivery tour (vehicle), respectively. At most, one size  $m$  and one type  $t$  of MFC can be built at each location  $l$  (Constraints (4)). Constraints (5) ensure that each tour starts at an MFC. The flow constraint in Eq. (6) defines that each location visited is also left again. According to Constraints (7), an MFC and a customer are on the same tour only if a customer is assigned to that MFC. Each tour starts at time 0 at the MFC (Constraints (8)). The arrival times at each customer location, i.e., the service start time at each customer, are then set by Constraints (9), where  $2D$  is used as “big M”. Constraints (10) calculate the lateness of a delivery at a customer location if the promised lead time is exceeded. Constraints (11) limit the picking volume assigned to an MFC to the available picking capacities and ensure that customers can only be assigned to activated MFCs. Constraints (12) limit the transportation volume



assigned to a vehicle such that the vehicle capacity  $Q$  is respected. Constraints (13) – (17) define the variable domains.

## 4.2 Solution approach

The CLRP\_MD is an LRP variant and a generalization of the CVRP (capacitated vehicle routing problem). It thus belongs to the class of NP-hard optimization problems (see, e.g., Toth and Vigo 2014). The interrelation of location selection, location type and size, customer assignment, and routing decisions drives the problem complexity of the CLRP\_MD. This requires solving a multiplicity of VRPs for changing customer assignments and location decisions (number, picking capacities, and locations of MFCs). Heuristic approaches are, therefore, needed to obtain insights into industry-relevant problem sizes of several thousand customers and a realistically large set of possible MFC locations. Nagy and Salhi (2007), Prodhon and Prins (2014), and Schneider and Drexl (2017) review LRP variants and show that sequential approaches are efficiently applied to LRPs. Sequential approaches solve the clustering and routing first and then allocate clusters to facilities or vice versa. Transferred to our problem, the latter case would imply that we first allocate customers to MFCs and thus determine the size, type, and location of MFCs before solving the resulting routing problems. This approach consequently neglects the fast delivery time aspiration in quick commerce as well as the major impact of transportation costs when deciding on locations and their sizes/types. The fast deliveries are, however, a major cost driver and essential for the fulfillment of the core service promises. An alternative and well-established approach is the approximation of transportation costs as input for the location and customer assignment (see, e.g., Barreto et al. 2007; Lopes et al. 2008). Quick commerce services have ambitious lead times, and small vehicles are applied, resulting in limited transportation capacity. These restrictions lead to short delivery tours with only a handful of customers on a tour. The limited tour lengths and sizes are advantageous when clustering customers to tours and enable an exact evaluation of transportation costs by solving a travelling salesman problem (TSP) for each cluster built. Figure 2 summarizes the proposed approach based on a CFRS methodology that approximates actual transportation and lateness costs for the final MFC selection and customer assignment. We first adapt the agglomerative hierarchical clustering (AHC) algorithm to our problem setting

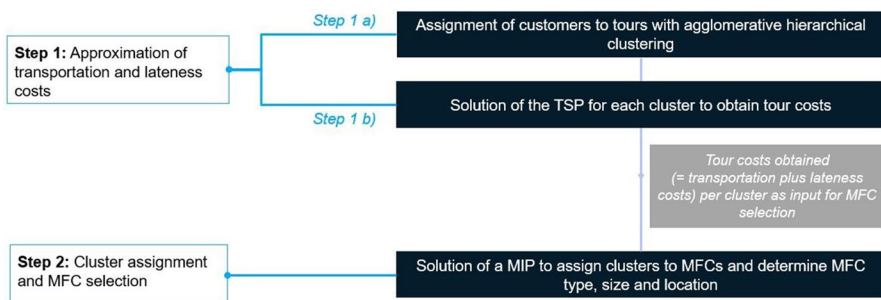


Fig. 2 Structure of the cluster-first-route-second approach

and subsequently solve the resulting TSPs for each cluster exactly (see Step 1: Approximation of transportation and lateness costs). Once we have obtained the transportation and lateness costs for all tours, we solve the customer assignment and location selection using an MIP formulation (see Step 2: Cluster assignment and MFC selection).

*Step 1: Approximation of transportation and lateness costs*

**Step 1a: Assignment of customers to tours with AHC.** The clustering of customers is a central aspect of our heuristic as it determines potential delivery tours. In our problem context, we need to cluster many customers from the entire delivery region into many small clusters due to tight lead times and capacity limits. We apply an AHC with a group average as a proximity measure for the clustering of customers. Barreto et al. (2007) show that the AHC constitutes an efficient approach for solving the clustering within LRPs. The authors analyze different clustering algorithms for LRPs and show that the AHC performs best based on standard LRP instances from Gaskell (1967), Christofides and Eilon (1969), Or (1976), Perl and Daskin (1985), Min et al. (1992), and Daskin (1997). In general, AHC starts with every customer as its own cluster and then merges clusters sequentially based on the proximity measure defined (Johnson 1967). In our application, the proximity measure calculates the average travel times between two clusters and defines the sequence in which promising clusters are merged (Barreto et al. 2007). We define  $C_a$  and  $C_b$  as two customer clusters and  $d_{ij}$  as the travel time between two customers. The group's average travel time is then defined as

$$\delta(C_a, C_b) := \frac{\sum_{i \in C_a} \sum_{j \in C_b} d_{ij}}{|C_a| \cdot |C_b|}. \quad (18)$$

Please note that MFCs are not yet considered for customer clustering. We further tailor the AHC to fit the requirements of quick commerce and thus implement four additional conditions for the clustering: inner-cluster travel times, travel times between customers and the next MFC location, maximum travel distance per customer to MFCs, and maximum vehicle capacity. The first clustering restriction reflects the aspired short lead times that apply in quick commerce. We exclude customer pairs with long travel times by limiting the possible driving time within a cluster with a maximum travel time between any two customers (denoted by  $\mu$ ). All customer pairs with  $d_{ij} > \mu$  cannot be merged in one cluster. Second, we additionally limit the number of customers within a cluster to limit the total tour duration, thus tightening the search for feasible clusters. Using the average travel time from customers to their nearest potential MFC location ( $\bar{\Delta}_1$ ), the average travel time between customers ( $\bar{\Delta}_2$ ) that can be merged (i.e., all customers  $i, j$  with  $d_{ij} \leq \mu$ ), and the average service time of customers ( $\bar{s}$ ), we define the maximum number of customers in a cluster as

$$C^{\max} := \left\lfloor \frac{D - \bar{\Delta}_1}{\bar{\Delta}_2 + \bar{s}} \right\rfloor. \quad (19)$$

Third, we need to ensure that each cluster created is feasible concerning the maximum travel distance from MFCs to individual customers. The maximum travel distance defines the delivery areas and may differ from the respective travel times, i.e.,

customer reachable within the lead time may still not be allocated to an MFC due to customer density or service areas defined. Considering the maximum travel distance per MFC, we only allow the clustering of customers that can be assigned to at least one joint MFC. Using the sets  $L_i$  reflecting the possible MFC assignments for each customer  $i$ , we define that two customer clusters  $C_a, C_b$  can only be merged if  $\bigcap_{i \in C_a \cup C_b} L_i \neq \emptyset$ . Finally, we can only add additional customers to a cluster as long as the vehicle capacity  $Q$  is not exceeded, i.e., Constraints (12) have to be respected. The vehicle capacity restricts the tour size if the lead times allow the clustering of multiple customers and reflects the limited capacities of small delivery vehicles used in quick commerce.

The adapted AHC, therefore, merges clusters according to the proximity measure Eq. (18) and adherence to the four additional criteria defined. Algorithm 1 illustrates the adapted AHC algorithm. Applying the AHC, all customers  $i \in C$  are assigned to customer clusters  $C_k$  ( $C_k \subset C$ ), where  $k \in K$  represents the set of clusters created. Each customer cluster  $C_k$  created will be the basis for the routing and customer allocation in the next steps.

#### Algorithm 1 Adapted AHC procedure

- 
- 1: Objective: Sequentially merge customers into clusters based on the proximity measure  $\delta$
  - 2: Input:  $C, d_{ij}, \mu, C^{\max}, Q, L_i \forall i \in C$
  - 3: Initiate each customer  $i \in C$  as its own cluster  $k$
  - 4: Calculate the proximity measure  $\delta$  between all pairs  $C_a, C_b$  of all existing clusters  $K$  using Eq. (18)
  - 5: Merge two clusters to a new cluster  $k$  that have the lowest proximity with  $\min\{\delta(C_a, C_b)\}$  while adhering to the maximum travel time between customers in the cluster  $\mu$ , the maximum number of customers per tour  $C^{\max}$ , the maximum travel distance between individual customers to MFCs ( $\bigcap_{i \in C_a \cup C_b} L_i \neq \emptyset$ ), and the vehicle capacity  $Q$
  - 6: Iterate steps 4 and 5 until no more merging is possible and save the clusters  $K$  obtained
  - 7: Return all clusters  $C_k, k \in K$
- 

**Step 1b: Solution of the TSP for each cluster to obtain tour costs.** In this subsequent step, we calculate the corresponding tour costs of transportation and lateness costs by solving the TSP for each cluster  $C_k$  obtained by the AHC. As customers and respective clusters have not yet been assigned to any specific MFC, the TSP is solved for each possible cluster-MFC combination, i.e., each TSP starts from the respective MFC, travels to the customers and returns to the MFC. The TSP is formulated as a dynamic program (see Held and Karp 1962) as it generally performs well for small instances (Applegate et al. 2011). Despite solving up to  $|K| \times |L|$  TSPs, we show in the numerical analysis that the TSPs can be solved optimally in a short computation time due to the small cluster sizes. The TSP minimizes the tour costs  $c_{lk}^{\text{tour}}$  for each cluster  $C_k$  supplied from MFC location  $l$ . These are calculated by

$$c_{kl}^{\text{tour}} = \sum_{i \in C_k \cup \{l\}} \sum_{j \in C_k \cup \{l\}} c_{ij}^{\text{trans}} \cdot x_{ij} + \sum_{i \in C_k} c_i^{\text{late}} \cdot r_i \quad \forall k \in K, l \in L \quad (20)$$

Here,  $x_{ij}$  is the reduced routing variable, indicating whether location  $j$  is visited after location  $i$  within one cluster. At the end of Step 1, we obtain tour costs for each customer cluster created and for each possible MFC as a starting point. We use the tour

costs obtained  $c_{kl}^{\text{tour}}$  as an approximation for the actual transportation and lateness costs of the CLRP\_MD and as input for Step 2.

*Step 2: Solution of a MIP to assign clusters to MFCs and determine MFC type, size and location.* Using the clusters and tour costs from Step 1 reduces the CLRP\_MD to an assignment and selection problem that assigns customer clusters to MFCs and selects the optimal locations, sizes and types of MFCs to be opened. As we consider customer clusters instead of individual customers in this step, we calculate the order processing costs ( $c_{klmt}^{\text{proc}}$ ) per cluster  $k$ , MFC location, size, and type by  $c_{klmt}^{\text{proc}} = \sum_{i \in C_k} c_{ilmt}^{\text{proc}}$ . The cumulative picking volume per cluster is defined by  $p_k = \sum_{i \in C_k} p_i$ . Moreover, we introduce  $z_{klmt}$  as a binary decision variable, indicating the assignment of clusters  $k$  to MFC location  $l$  with size  $m$  and type  $t$ , and the binary variable  $y_{lmt}$  to indicate whether location  $l$  will be chosen to open an MFC at location  $l$  with size  $m$  and type  $t$ . The assignment and selection problem is then formulated as follows:

$$\text{Min } FC^*(y, z) = \sum_{l \in L} \sum_{m \in M} \sum_{t \in T} (c_{lmt}^{\text{fix}} + c_{lmt}^{\text{rep}}) \cdot y_{lmt} + \sum_{k \in K} \sum_{l \in L} \sum_{m \in M} \sum_{t \in T} (c_{klmt}^{\text{proc}} + c_{kl}^{\text{tour}}) \cdot z_{klmt} \quad (21)$$

subject to

$$\sum_{l \in L} \sum_{m \in M} \sum_{t \in T} z_{klmt} = 1 \quad \forall k \in K \quad (22)$$

$$\sum_{k \in K} \sum_{m \in M} \sum_{t \in T} p_k \cdot z_{klmt} \leq \sum_{m \in M} \sum_{t \in T} S_{mt} \cdot y_{lmt} \quad \forall l \in L \quad (23)$$

$$\sum_{m \in M} \sum_{t \in T} y_{lmt} \leq 1 \quad \forall l \in L \quad (24)$$

$$y_{lmt} \in \{0, 1\} \quad \forall l \in L, m \in M, t \in T \quad (25)$$

$$z_{klmt} \in \{0, 1\} \quad \forall k \in K, l \in L, m \in M, t \in T \quad (26)$$

The Objective Function (21) minimizes total fulfillment costs based on the selection of MFC locations, sizes, and types and the cluster assignment to MFC locations. The first term sums up the fixed setup and replenishment costs of MFCs opened, while the second term sums up total order processing and tour costs. Constraints (22) denote that each cluster has to be assigned to exactly one MCF, whereas Constraints (23) ensure that the total picking capacity of each MFC is respected. Constraints (24) define that one MFC type and size is selected at most at each MFC location. Finally, the variable domains are defined by Constraints (25) and (26). The MIP returns the final solution of the CLRP\_MD.

## 5 Numerical analysis

This section presents the numerical analyses. First, we lay out the test data applied in Sect. 5.1. Section 5.2 analyzes the performance of our approach by comparing it to the optimal solution of the CLRP\_MD for small problem sizes. We then provide managerial insights on the strategic network design for quick commerce in Sect. 5.3.

### 5.1 Generation of test data

Customer locations  $C$  are generated based on existing buildings in OpenStreet-Map (2022) in Munich. We chose an area in the most populated region of around  $14 \times 23 \text{ km}^2$  for customer and MFC locations. Using this approach, there are 83,551 potential customer locations to choose from. The travel time  $d_{ij}$  between all locations amounts to up to 107 min with a median of 25 min. We create various customer samples from this customer database to obtain multiple test instances and analyze different settings. Each sample is formed using a randomly selected subset of customers. We allow up to 30 different MFC locations. They are evenly distributed within the delivery area in the form of a rectangular grid laid over the delivery area. For 30 MFCs, the distance between the potential locations is roughly 3.5 km for two neighboring MFCs. Three MFC sizes  $m \in M$  (small, medium and large) with  $250 \text{ m}^2$ ,  $400 \text{ m}^2$ , and  $750 \text{ m}^2$  and two types  $t \in T$  (manual and automated) are possible for each potential location. To streamline the analysis, we assume that automation applies only to larger MFCs. This reduces the possible size-type combinations to three distinct options (small and medium MFCs with manual picking, large automated MFCs). The type and size determine the picking capacity  $S_{mt}$  (Wulfraat 2019), set at 1000, 2000, and 5000 units, respectively.

The fixed setup costs  $c_{lmt}^{\text{fix}}$  are calculated based on average market rental costs per period for commercial buildings in Munich (in 2022), with a 30% surplus as costs for utilities for manual picking and 50% for automated picking (see Statista 2023a, extrapolated with 2023 Q1 market listings in Munich). We assume identical costs across all locations for each MFC size-type combination, which then results in  $c_{lmt}^{\text{fix}}$  of 650€, 1040€, and 1875€. The order processing costs  $c_{ilmt}^{\text{proc}}$  are based on insights from picking in online retailing (see, e.g., Boysen et al. 2019; Dethlefs et al. 2022). They mainly depend on the basket size of a customer  $i$  and the travel distance of the picker in the MFC. The actual basket sizes range between five and 15 items, with an average of 10 items (Statista 2022b). We determine the order processing costs, following Dethlefs et al. (2022), who estimated the picking and packing costs with manual picking for small depots. Using their approach, an order size of 10 items results in  $c_{ilmt}^{\text{proc}} = 1.14\text{€}$  for small MFCs with manual picking as an example. The costs are then obtained for all customer order sizes accordingly. Furthermore, we apply identical picking costs for the two manual MFC sizes as we assume a balancing effect for longer picking routes and possible layout optimization (less congestion) for medium-sized MFCs. The processing costs for the large, automated MFCs are 13% lower due to a higher throughput rate (see Caputo and Pelagagge 2006; Wulfraat 2019, adjusted to sophisticated equipment). Finally, identical processing

costs are applied across locations, i.e.,  $c_{ilmt}^{\text{proc}}$  ultimately differ by MFC type and size as well as order size. The replenishment costs  $c_{ilmt}^{\text{rep}}$  are dependent on the distance from the CDC and the volume processed in the MFC. The CDC is located 30 km northeast of the geographical center of the delivery region, and its location is identical throughout all test instances. Each MFC is replenished frequently with full truck loads (FTLs). We apply a cost factor of 3.23€/km for transport from the CDC to the MFC (see Transport Intelligence Limited 2022). The size of the MFC determines the replenishment frequency and volume, which results in 20% and 30% lower costs for small and medium-sized MFCs compared to automated MFCs. We assume in our study that the impact of automation solutions is restricted to the picking process and that replenishment costs do not depend on the MFC type. The transportation costs  $c_{ilmt}^{\text{trans}}$  are set at 1.00€/km and calculated based on minimum wage (incl. non-wage costs) as well as depreciation and energy usage of vehicles. The retailer applies a homogeneous fleet across MFCs. As identical transportation boxes for all customers are applied, we vary the vehicle capacity  $Q$  between 1 and 5 (see capacity of vehicles associated with small depots in Dethlefs et al. 2022). If not stated otherwise, we set the aspired lead time  $D$  at 60min and the service time  $s_i$  at 3 min per customer stop (see, e.g., Ulmer et al. 2022). The travel times are determined assuming an average vehicle speed of 15 km/h for electric bicycles in urban traffic (see, e.g., Fishman and Cherry 2016). Lastly, the penalty costs  $c_i^{\text{late}}$  are set prohibitively high at 1.00€/min for each minute of delay.

We use  $\mu = 6$  min as a basic value for the clustering algorithm and allow each customer to be allocated to each MFC, i.e.,  $L_i = L, \forall i \in C$ . The average travel times from MFCs to a potential customer ( $\Delta_1$ ) and between customers in a potential cluster ( $\Delta_2$ ) are determined for each instance upfront.  $\Delta_1$  ranges between 8.7 and 8.8 min and  $\Delta_2$  between 3.6 and 4.0 min. If not stated otherwise, we apply the data set specified above in all our experiments. The heuristic is implemented in Python 3.8.14, using Gurobi 10.0.0. as the solver for the MIP. The tests were run on an Intel(R) Core(TM) i7-1185G7 CPU @ 3.00GHz.

## 5.2 Efficiency and runtime analysis of the heuristic

*Comparison to exact approach.* The NP-hard CLRP\_MD (see (1) to (17)) can only be solved exactly for small data instances using a commercial solver. The number of locations, types and sizes of MFCs, and the number of customers drive the combinatorial complexity. We, therefore, solve the CLRP\_MD with varying problem sizes for a reduced delivery area of  $4 \times 4$  km<sup>2</sup>. To obtain different costs per MFC in this small area, we apply  $c_{ilmt}^{\text{rep}}$  2.5 times higher than in the base case defined above. Table 4 summarizes the average runtime and solution value of 10 instances for each problem size. We set a time limit of 3 h for the Solver for all tests. Our heuristic provides solutions close to optimality within seconds. In contrast, an exact solution already requires more than 20 min for ten customers and ten MFCs and two size/technology options as an example.

The results show that the solution gap is less than 0.8% across the test instances that could be solved to optimality with regard to the MIP gap of 0.01%. This shows that our

**Table 4** Comparison with exact solution of the CLRP\_MD

Problem size			Instances solved to optimality with MIP gap = 0.01%					Instances solved to optimality with MIP gap = 1% <sup>1</sup>					
L	M ∪ T	C	Inst. <sup>2</sup>	MFC <sup>3</sup>	Runtime [sec.]		Cost delta, in % of Gurobi <sup>5</sup>	Inst. <sup>2</sup>	MFC <sup>3</sup>	Runtime [sec.]		LB gap <sup>4</sup>	Cost delta, in % of Gurobi <sup>5</sup>
					Heur	Gurobi				Heur	Gurobi		
5	3	5	10	10	0.03	0.20	0.12%	—	—	—	—	—	—
10	2	5	10	10	0.03	0.43	0.15%	—	—	—	—	—	—
5	3	10	10	9	0.04	468	0.48%	—	—	—	—	—	—
10	2	10	10	9	0.03	1274	0.48%	—	—	—	—	—	—
5	3	15	2	2	0.09	742	0.71%	5	4	0.09	745	0.97%	0.42%
10	2	15	2	2	0.07	3914	0.80%	4	4	0.06	3.479	0.93%	0.42%
5	3	20	0	—	—	—	—	0	0	0.06	—	—	—
10	2	20	0	—	—	—	—	0	0	0.04	—	—	—

<sup>1</sup>Instances that could not be solved with a MIP gap of 0.01% within 10,800 sec runtime have been additionally solved with a MIP gap of 1.0%<sup>2</sup>Number of data instances that could be solved to optimality (with regard to the MIP gap)<sup>3</sup>Number of instances where heuristic and Gurobi resulted in identical MFC locations, types, and sizes<sup>4</sup>Gap between best MIP solution after 10,800 sec runtime and lower bound provided by Gurobi<sup>5</sup>Gap between heuristic and MIP (Gurobi), calculated as obj. value obtained by heuristic divided by obj. value obtained by Gurobi



**Table 5** Runtime of the heuristic with  $|L| = 10$  and  $|M \cup T| = 3$ , average of 10 instances, in sec

C	5000				10,000				15,000			
	2	3	4	5	2	3	4	5	2	3	4	5
$C^{max}$												
Step 1a	1	1	1	1	4	3	3	3	9	8	8	8
Step 1b	2	2	2	3	5	4	5	5	9	7	7	9
Step 2	137	80	65	68	161	150	133	129	373	231	225	213
Total	140	83	68	71	170	158	141	138	391	245	239	230
Min. total	127	71	36	39	72	92	106	59	274	179	127	118
Max. total	150	98	92	90	301	209	192	238	571	300	336	288

approach is able to find a good approximation for the transportation costs and provides a solid basis for the subsequent MFC selection. Of course, the clustering limits the tour building due to a fixed assignment of customers to clusters and tends to create more and smaller clusters, i.e., delivery tours. As small tours are realistic due to the time difference of orders, our approach represents a viable approach for an actual industry application.

*Runtime analysis for larger problems.* We analyze the runtime performance of our approach for larger problems. Our algorithm consists of three steps (1a: Clustering, 1b: TSP, 2: MIP for cluster assignment). We consequently depict the runtime of each solution step. We vary the number of customers, locations, MFC types/sizes and set a maximum number of customers per cluster  $C^{max}$  for our experiments. We define cluster sizes for this test externally, as the number of tours is expected to impact the runtimes. Table 5 summarizes the tests.

Step 2 (solving the assignment MIP) is the main driver for computation time in our approach. Steps 1a (clustering) and 1b (solving the TSP) are not computational bottlenecks. We find an optimal solution for the TSPs in a short time as we only need to deal with small TSPs of up to five customers. The number of tours naturally increases with smaller tours (imposed by a smaller  $C^{max}$ ). This drives the runtime of Step 2. Yet we are considering a strategic problem, and runtimes of less than 400 sec for realistic problem sizes are acceptable.

### 5.3 Managerial insights

We conduct further experiments within an application for a large city (see data above) to investigate the impact of central problem parameters on the MFC network, on-time delivery, tour length, and fulfillment costs. We first vary the lead time promise as the short time from order placement to delivery is the novelty of this business. Second, we analyze the impact of increasing bundling on tours, i.e., increasing vehicle capacities, as this is a main driver for transportation costs. Third, we study the impact of the network setup by comparing a network with fewer large MFCs to a network with many small MFCs based on the corresponding picking and setup costs. Fourth, to tackle the higher fixed costs together with the possible scarcity of city center locations for MFCs, we look into the option of building MFCs in the outer rim of the delivery area instead of opening them in the middle of it.

### 5.3.1 Analysis of delivery requirements in quick commerce

Quick commerce retailers distinguish themselves by their lead time promises. These are pivotal for the required MFC network, the daily operations, and the cost structure. We, therefore, analyze two central aspects of the delivery to obtain managerial insights on the impact of (i) *lead times promises* and (ii) *bundling options on tours*.

(i) *Impact of lead time promises.* The maximum lead times  $D$  determine the possible routing options. Shorter lead times limit the allowed route duration and increase possible lateness costs for deliveries. We, therefore, analyze how changes from  $D = 15$  to  $D = 60$  min impact the network, on-time delivery and costs, with  $D = 15$  min as the baseline. This analysis assumes a sufficiently large vehicle capacity  $Q$  to focus on the lead time effects. The lead time also impacts the maximum number of customers within a cluster in our solution approach (see determination of  $C^{max}$  with Eq. 19).

Figures 3 and 4 visualize the changes in MFC networks and in customer assignments for  $D = 15$  and  $D = 60$  min. Each group of customers in the same grey gradient is assigned to the MFC located in the containing area. The example shows that a longer lead time requires fewer MFCs (see, e.g., less populated area in the top right).

Table 6 shows that aiming for  $D = 60$  min instead of 15 min would require only five instead of nine MFCs with the smaller customer base ( $|C| = 5,000$ ). With a larger customer base, expanding  $D$  also results in a lower required number of MFCs. When comparing the different customer sizes  $|C|$  for the identical lead time  $D$ , it becomes obvious that mainly the MFC sizes and their utilization increases (see, e.g.,  $|C = 5,000|$  and  $D = 15$  vs.  $|C = 15,000|$  and  $D = 15$ ). The lead time expansion

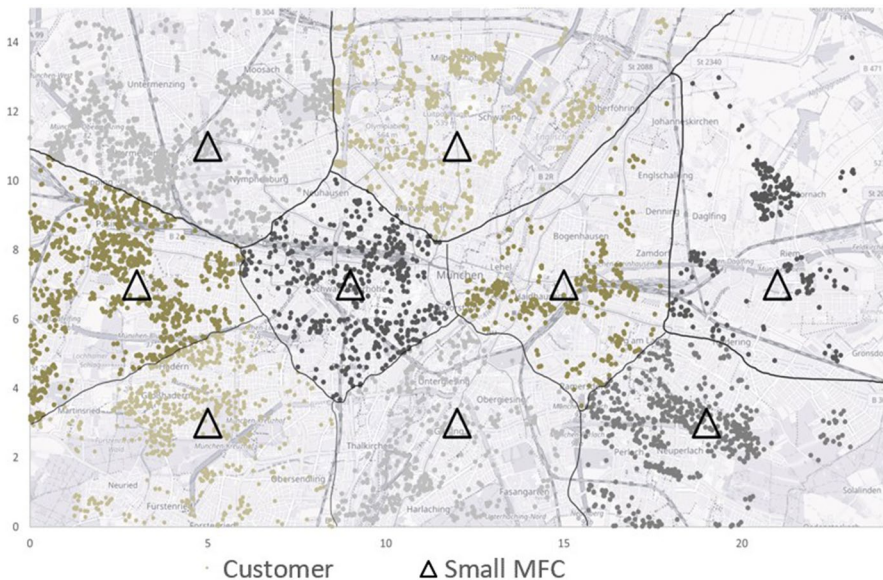
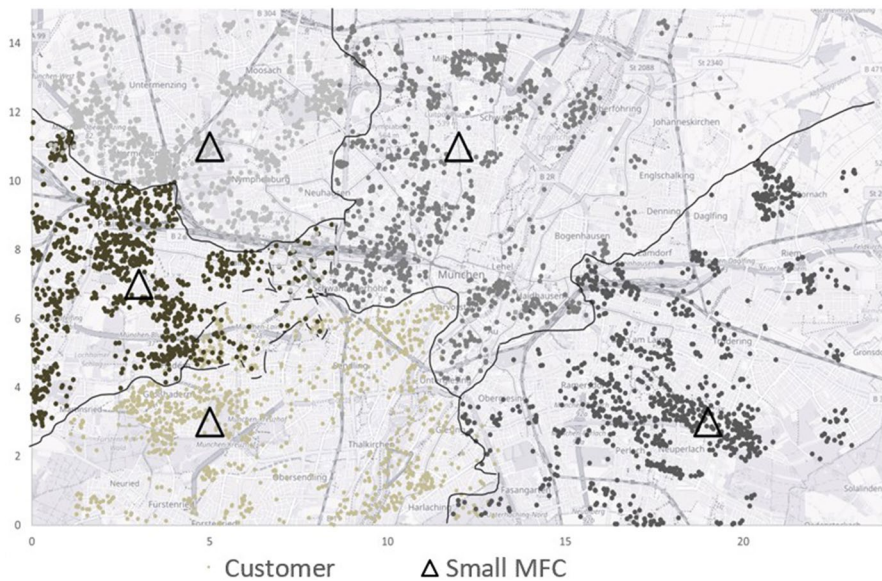


Fig. 3 MFC network for  $D = 15$  min, with  $|C| = 5,000$ ,  $|L| = 10$ , example instance



**Fig. 4** MFC network for  $D = 60$  min, with  $|C| = 5,000$ ,  $|L| = 10$ , example instance

from 15 to 60 min allows cost savings of 56–57% across all instances. High savings in transportation costs mainly achieve this. In all cases, lateness costs are minor as the on-time delivery is still achieved through opening a sufficient number of MFCs.

(ii) *Bundling options in tours.* The bundling options of different customers on a tour are expected to drive the transportation costs. Furthermore, as delivery vehicles (e.g., cargo bikes) in quick commerce are generally small, retailers must deal with their limited transportation capacity. We analyze the bundling options on a tour by investigating the impact of transporting one to five orders per vehicle per tour, i.e., limiting the vehicle capacity accordingly. To assess the impact of vehicle capacity  $Q$  on costs and network, we set the maximum lead time at  $D = 60$ . We apply direct deliveries with  $Q = 1$  as the benchmark in this analysis. Table 7 summarizes the impact of varying bundling options with different vehicle capacities  $Q$ .

We observe similar effects for varying the bundling options as for the lead times analysis. This can be attributed to the maximum tour sizes that can be obtained in both experiments. Larger vehicle capacities  $Q$  and larger lead times  $D$  are necessary to obtain transportation costs savings. A larger vehicle capacity leads to a higher bundling of customers and results in a significant reduction of total fulfillment costs by up to 51%. Increasing the capacity from one to two orders per tour, the total fulfillment costs can already be reduced by 26 to 29%. Together with the respective MFC choices, the share of transportation costs decreases by around 30 percentage points. The analysis further highlights that limiting  $Q$  restricts tour sizes and hence also limits late deliveries. Across all instances, almost no lateness costs apply, and customers receive their orders with a median of less than 33 min. The median arrival time for direct deliveries is around nine minutes, and with two orders per tour, this value increases to

**Table 6** Impact of increasing delivery lead times  $D$ , with  $|L| = 10$ , average of 10 instances

$ C $	5000					10,000					15,000				
	15	30	45	60		15	30	45	60		15	30	45	60	
FC <sup>1</sup>	0%	– 38%	– 50%	– 56%		0%	– 40%	– 51%	– 56%		0%	– 41%	– 52%	– 57%	
Set/Repl. <sup>2</sup>	18%	24%	25%	23%		12%	19%	22%	23%		11%	17%	20%	22%	
Proces. <sup>2</sup>	16%	26%	32%	36%		18%	30%	36%	40%		18%	30%	36%	40%	
Transp. <sup>2</sup>	63%	49%	43%	41%		68%	51%	42%	37%		70%	53%	43%	38%	
Late <sup>2</sup>	2%	1%	< 1%	< 1%		2%	< 1%	< 1%	< 1%		2%	1%	< 1%	< 1%	
No. of tours	5000	1695	1034	750		10,000	3363	2034	1497		15,000	5030	3034	2181	
Arriv. time <sup>3</sup>	8.69	21.43	30.80	39.57		8.70	20.58	30.86	38.35		8.66	20.64	29.73	38.16	
Long. tour <sup>4</sup>	55.35	57.29	70.70	74.13		48.25	60.34	65.52	73.58		47.51	61.78	63.78	60.00	
Small MFC <sup>5</sup>	9.0	7.1	6.0	5.0		8.0	7.0	6.6	6.0		4.0	3.9	4.0	4.4	
	56%	70%	83%	100%		76%	86%	92%	100%		70%	93%	92%	98%	
Med. MFC <sup>5</sup>	0.0	0.0	0.0	0.0		2.0	2.0	2.0	2.0		5.0	4.1	4.0	3.3	
						98%	99%	100%	100%		89%	95%	96%	99%	
Large MFC <sup>5</sup>	0.0	0.0	0.0	0.0		0.0	0.0	0.0	0.0		1.0	1.0	1.0	1.0	
											66%	72%	73%	82%	

<sup>1</sup>Change of total fulfillment costs (FC) in comparison to  $D = 15$  for each  $|C|$ , in %<sup>2</sup>Share of fixed setup and replenishment costs, order processing costs, transportation costs and lateness costs as % of total fulfillment costs<sup>3</sup>Median arrival time at the last customer on a tour (incl. travel times to all customers and service times at previous customers), in min<sup>4</sup>Longest total travel time across all tours, including return travel from last customer to the MFC, in min<sup>5</sup>Average number of the MFCs (first line) and its average utilization (second line)

**Table 7** Impact of order bundling with increasing vehicle capacity, with  $|L| = 10$  and  $D = 60$ , average of 10 instances

IC <sup>1</sup>	5000					10,000					15,000				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
FC <sup>1</sup>	–	– 26%	– 38%	– 45%	– 49%	–	– 28%	– 39%	– 46%	– 50%	–	– 29%	– 40%	– 47%	– 51%
Set/Repl. <sup>2</sup>	19%	23%	20%	23%	25%	12%	16%	19%	21%	21%	11%	14%	17%	19%	20%
Proces. <sup>2</sup>	17%	22%	27%	30%	32%	18%	25%	30%	33%	36%	18%	25%	30%	34%	36%
Transp. <sup>2</sup>	65%	55%	53%	47%	43%	70%	59%	51%	46%	43%	71%	61%	53%	48%	43%
Late <sup>2</sup>	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%
No. of tours	5000	2520	1695	1282	1034	10,000	5022	3363	2531	2034	15,000	7523	5030	3784	3034
Arriv. time <sup>3</sup>	8.69	15.43	21.95	26.56	30.80	9.48	15.24	20.56	25.30	32.68	8.56	15.19	20.63	25.40	29.74
Long. tour <sup>4</sup>	56.92	57.77	70.20	70.58	70.70	54.14	66.16	66.56	68.62	72.56	47.51	60.28	61.17	61.25	61.58
Small MFC <sup>5</sup>	9.0	8.0	6.0	6.0	6.0	7.9	7.0	7.0	7.0	6.2	4.0	3.6	4.0	4.0	4.0
	56%	63%	83%	83%	83%	77%	86%	86%	86%	97%	69%	91%	93%	93%	92%
Med. MFC <sup>5</sup>	0.0	0.0	0.0	0.0	0.0	2.0	2.0	2.0	2.0	2.0	5.0	4.3	4.0	4.0	4.0
						98%	99%	99%	99%	100%	89%	95%	96%	96%	96%
Large MFC <sup>5</sup>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0
											66%	70%	72%	73%	73%

<sup>1</sup>Change of total fulfillment costs (FC) in comparison to direct deliveries (i.e.  $Q = 1$ ), in %

<sup>2</sup>Share of fixed setup and replenishment costs, order processing costs, transportation costs and lateness costs as % of total fulfillment costs

<sup>3</sup>Median arrival time at the last customer on a tour (incl. travel times to all customers and service times at previous customers), in min

<sup>4</sup>Longest total travel time across all tours, including return travel from last customer to the MFC, in min

<sup>5</sup>Average number of the MFCs (first line) and its average utilization (second line)

around 15min and remains below 33 min even for tours with up to five customer stops. However, the short lead times are achieved by establishing up to nine MFCs. Especially in instances with 5000 customers, MFCs are opened with a low average utilization of 56%. In general, a decreasing bundling of customers requires the establishment of more MFCs and results in a lower utilization per MFC. The bundling consequently contributes to the efficiency of both small and large MFCs.

### 5.3.2 Analysis on MFC network options

The number, size and type of MFCs used and their setting (i.e., size and type) is the central element of quick commerce distribution. The available MFC network defines the possible service offerings to customers and the delivery options. In practice, quick commerce retailers have to work with the existing locations and related setting possibilities that are available on the market. This setting may significantly impact the fulfillment costs for the retailers and their service offerings. We subsequently analyze two different scenarios for the available MFC infrastructure by looking on (i) different MFC sizes (small vs. large) and (ii) general locations (center vs. outskirt).

(i) *Small vs. large MFCs.* A central aspect of quick commerce distribution networks is the question of using fewer but larger automated MFCs or establishing a distributed network of many small and manually operated MFCs to service the delivery region. Automated MFCs have higher fixed setup costs that enable lower order processing costs, while the opposite applies to small, manual MFCs. To compensate for the higher fixed costs, automated MFCs may only be cost-efficient within larger delivery regions. However, this implies as well higher transportation distances and costs. To analyze the impact of the MFC size and type setting on costs and network, we construct two contrary MFC options to highlight the different effects. The first option uses only small, manual MFCs and constitutes the benchmark (i.e., no other MFC types and sizes are available). The second option is the implementation of one large, automated MFC that serves the entire delivery region and has sufficient capacity to process all customers. The required large MFC sizes impact the fixed setup costs  $c_{lmt}^{fix}$ , which are set at 3750€, and 5625€, for 10,000 and 15,000 customers, respectively. We allow up to 30 locations and restrict  $Q = 5$ . We allow  $D = 60$  to be not too restrictive on the possible network configuration (e.g., with  $D = 15$  min, the extreme scenario with one automated MFC results in extremely high lateness costs). All other parameters remain unchanged.

The results (Table 8) show that multiple smaller MFCs are imperative for quick commerce fulfillment. In other words, a network of distributed MFCs with manual order processing outperforms the use of a single large, centralized, automated MFC. The significant increase in transportation costs with only one central MFC counteracts the savings achieved through automated order processing. Installing a central MFC becomes even less attractive with an increasing customer base, as increased transportation costs to more customers further outweigh the savings resulting from automated order processing. This means that the fulfillment needs to be conducted with many small decentralized MFCs, even if the quick commerce retailers continue to grow. Many distributed smaller MFCs enable much faster deliveries, and the median arrival time to the last customer is

**Table 8** Impact of MFC types, with  $|L| = 30$ ,  $Q = 5$ ,  $D = 60$ , average of 10 instances

ICI	5000		10,000		15,000	
	Manual	Automated	Manual	Automated	Manual	Automated
Fulfillment costs	—	+ 36%	—	+ 48%	—	+ 55%
Setup/replen.costs <sup>2</sup>	21%	8%	24%	8%	25%	8%
Processing costs <sup>2</sup>	32%	21%	36%	21%	38%	21%
Transportation costs <sup>2</sup>	47%	66%	40%	66%	37%	66%
Lateness costs <sup>2</sup>	< 1%	4%	< 1%	5%	< 1%	5%
Median arrival time <sup>3</sup>	31.63	45.89	28.97	46.07	27.77	46.17
Longest tour duration <sup>4</sup>	76.64	127.80	62.05	130.05	47.82	130.14
No. of MFCs	5.0	1.0	10.5	1.0	15.9	1.0

<sup>1</sup>Manual, small MFCs; Automated, large MFC

<sup>2</sup>Share of fixed setup and replenishment costs, order processing costs, transportation costs and lateness costs as % of total fulfillment costs

<sup>3</sup>Median arrival time at the last customer on a tour (incl. travel times to all customers and service times at previous customers), in min

<sup>4</sup>Longest total travel time across all tours, including return travel to the MFC, in min

about 50% lower than with one centralized MFC. Although lead time promises can still be achieved with one central MFC, the tour duration increases significantly.

We create a second scenario to analyze the impact of order processing costs for MFC selection and a possible shift towards automated MFCs. We use the identical benchmark as above but allow multiple automated MFCs with a capacity of 5000 units (as in the base case) and conduct sensitivity analysis with decreasing order processing costs for the automated MFCs.

Table 9 shows that automated MFCs only provide cost benefits if processing costs are significantly lower. Furthermore, the threshold of integrating automated MFCs in the network becomes lower with an increasing customer base. This confirms the results of Table 7. Yet, we did not account for the potentially higher investment needed to achieve the lower processing costs. If additional setup costs are required to decrease the processing costs, automation becomes even less cost-efficient. In general, the results of our analyses indicate that automation may not yield substantial economic benefits for quick commerce fulfillment. The main reason is the comparatively low picking costs for small orders in manual MFCs. Additionally, multiple smaller MFCs distributed in the delivery area result in lower transportation costs compared to fewer automated MFCs. Furthermore, smaller non-automated locations may be easier to realize as real estate property in city centers for larger automated MFCs may be much more difficult to get. To summarize, quick commerce retailers should only consider investing in automation if two preconditions are met: The customer base is sufficiently large, and the processing costs are at least 40% lower compared to the current cost ratio between automated and manual MFCs. Even in those cases, the investment required for the automation technology still needs to be evaluated.



**Table 9** Impact of relative processing costs on MFC type selection, with  $|L| = 30, Q = 5, D = 60$ , average of 10 instances

IC1	5000			10,000			15,000		
	– 20%	– 40%	– 60%	– 80%	– 20%	– 40%	– 60%	– 80%	– 20%
Change of processing costs <sup>1</sup>									
Fulfillment costs <sup>2</sup>	0%	0%	– 1%	– 5%	0%	– 2%	– 7%	– 14%	– 1%
Median arrival time <sup>3</sup>	31.62	31.63	35.17	37.35	29.08	30.46	32.57	33.30	28.10
No. of manual MFCs	5.0	5.0	2.4	0.4	9.7	5.4	0.5	0.0	12.6
No. of automated MFCs	0.0	0.0	1.0	2.0	0.3	1.8	4.0	4.0	1.0

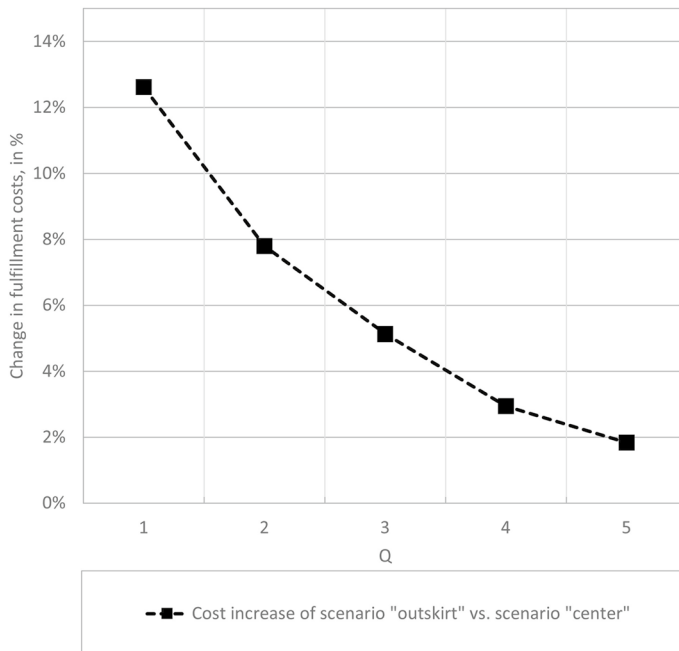
<sup>1</sup>Change of order processing costs in automated MFCs compared to base case; no change of costs in manual MFCs

<sup>2</sup>Change of total fulfillment costs (FC), in comparison to benchmark with only manual MFCs, in %

<sup>3</sup>Median arrival time at the last customer on a delivery tour (incl. travel times to and service times at previous customers), in min

(ii) *City center vs. outskirt area MFC network.* Quick commerce retailers operate in city centers to realize short lead times. They often use spaces of closed-down businesses in city centers (see, e.g., Butler 2021; tagesschau.de 2023). Additionally, MFCs can also be established within parts of larger commercial spaces previously used by department stores or similar. Suitable and various-sized spaces may hence become available. Nonetheless, if locations are available, rental prices in central areas are significantly more expensive compared to more remote areas. Taking into account these developments, we test the impact of locating MFCs only on the borders of the delivery region. We use the base case again as the benchmark and denote this scenario as “center”. We further apply a 25% reduction of fixed setup costs in a second scenario (denoted as “outskirt”) to reflect the lower rental prices in the outskirts of the city for comparison. There are ten possible MFC locations within a 1 km distance from the edge of the delivery area, again evenly distributed. Locations in the remaining area are not possible in the scenario “outskirt”. We test the impact of MFC locations again in dependence on  $Q$  to analyze the tour size impact on the MFC location setting.

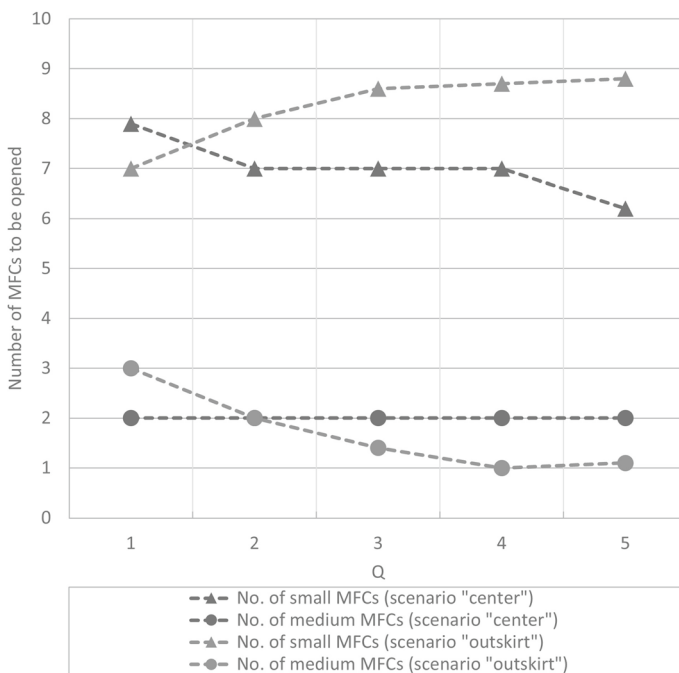
Our result shows that allowing MFCs only in the outskirt locations leads to higher total fulfillment costs. Figure 5 indicates the cost increase between the center and outskirt scenario. Despite lower setup and rental costs, it is always more expensive to locate MFCs in the outskirts. The fewer customers that can be bundled (i.e., decreasing  $Q$ ), the higher is the cost advantage of center locations.



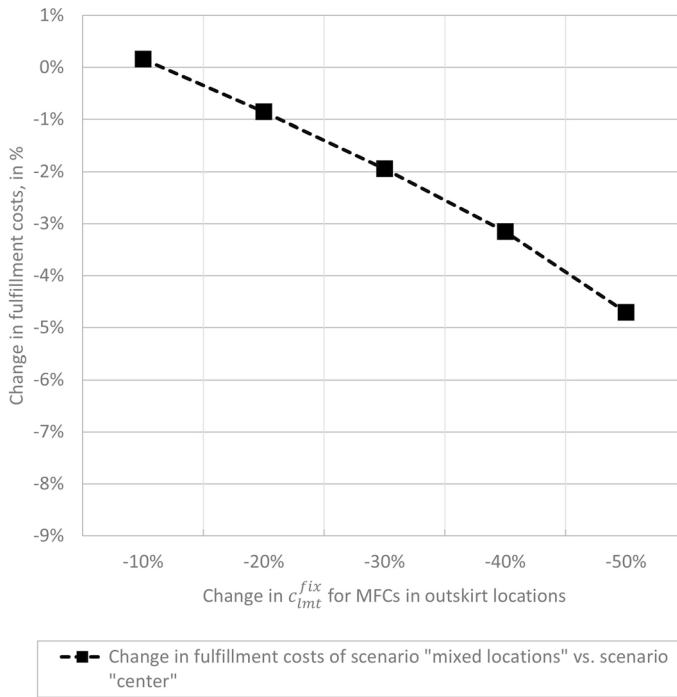
**Fig. 5** Change of fulfillment costs when MFCs can not be located in inner city, with  $|C| = 10,000$ ,  $|L| = 10$ ,  $D = 60$ , average of 10 instances

For details, please refer to Table 10 in the Appendix 1. The main reason lies in the significantly higher transportation costs, which offset lower fixed setup costs. However, the median arrival time is only about 1–2min higher when delivering not from the city center, thus enabling efficiently fast deliveries. This result can be attributed to two facts. Firstly, locations at the edge of the delivery area can still be within a reasonably limited distance to customers in densely populated cities. Secondly, opening more smaller MFCs helps to lower transportation costs. Large MFCs are not part of the solutions. The MFC size and location choices show that the preferable MFC setups are smaller with more of them overall when compared to the “center” scenario due to the more remote locations of MFCs as shown in Fig. 6. This setup becomes even more cost-efficient if customer orders are bundled together, thus lowering the average tour cost per individual customer order. To summarize, even with a 25% reduction of fixed setup costs in the outskirts, it is still more cost-efficient for the retailer to establish MFCs in central areas.

Following these insights on MFC locations, we extend the analysis by combining outskirt locations with locations at the very heart of the delivery region. This scenario is denoted as “mixed locations”. We include ten center locations (4.5 km from the edge of the delivery area) and ten outskirt locations for the MFCs (as above). This is compared to the “center” scenario.

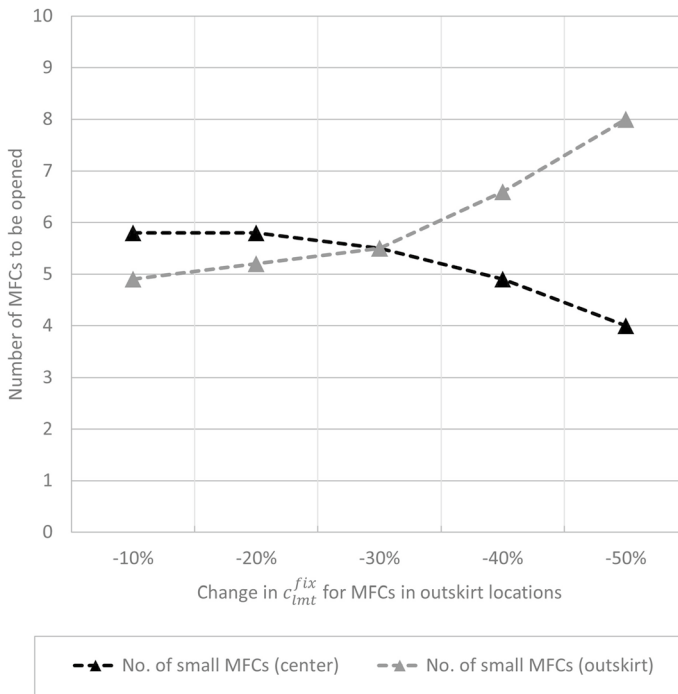


**Fig. 6** Number and Type of MFCs, with  $|C| = 10,000$ ,  $|L| = 10$ ,  $D = 60$ , average of 10 instances



**Fig. 7** Change of fulfillment costs, with  $|C| = 10,000$ ,  $|L| = 20$ ,  $D = 60$ ,  $Q = 5$ , average of 10 instances

Assuming 50% lower setup costs for MFCs in the outskirts decreases the total fulfillment costs by almost 5% without having to compromise on their lead time promises (see Fig. 7 and Appendix, Table 11). The underlying reasons are like the above. Outskirt locations have higher transportation distances that partially take away the savings in fixed costs for the setup. Figure 8 shows that with more than 30% lower fixed costs, the mix of MFCs also changes more from center locations to more outskirt locations. With the increase of more small MFCs, the customers still receive their orders within the promised lead time.



**Fig. 8** Number of MFCs for scenario “mixed locations”, with  $|C| = 10,000$ ,  $|L| = 20$ ,  $D = 60$ ,  $Q = 5$ , average of 10 instances

### 5.3.3 Summary of managerial insights

Our analyses show that creating the opportunity to serve several customers on one delivery tour through either longer lead time or more bundling options substantially decreases total fulfillment costs and the selection of MFC locations, sizes and numbers. The total fulfillment costs can already be reduced by one-third with as little as two stops per tour instead of direct deliveries to each customer. Furthermore, lead time promises impact the number of required MFCs to a large extent. Economies of scale are possible with larger lead times. As a result, an increase in lead time does not just reduce transportation costs, it also has an effect on the number of MFCs required. Larger and automated MFCs only become attractive and well utilized when larger tours can be built. This is only possible again with longer lead times and larger customer bundling. As transportation costs dominate, we can conclude that multiple smaller MFCs are more cost-efficient for quick commerce distribution. With regard to MFC locations, our findings demonstrate that building MFCs on the outskirts instead of the city center is not the best option. City center locations are required, again due to the dominant role of transportation costs. We summarize and highlight the three major managerial insights as follows:

- A lead time promise of 15 min is extremely costly to achieve. Just by targeting 30 min instead of 15 min, about 40% of the fulfillment costs can be saved.
- Bundling customers on tours and using vehicles with larger capacities reduces total fulfillment costs due to more efficient routes. This also impacts the MFC locations: When several orders can be delivered on one tour, fewer MFCs are required, resulting in a network with fewer but better utilized MFCs.
- Several small manual MFCs instead of one large automated MFC are key to planning a cost-efficient quick commerce distribution network. Having these small manual MFCs throughout the delivery area is essential in keeping transportation costs low.

## 6 Summary, limitations and future avenues of research

*Summary.* Our work is the first contribution to analyze the strategic network problem for quick commerce retailing. This paper investigates the impact of where, how many, how large, and what type of MFCs should be established in a quick commerce distribution network. We analyze the resulting total fulfillment costs and arrival times at customers. Our model and solution approach, based on a cluster-first-route-second method, are tailored to the requirements of quick commerce and enable efficient decision support.

*Limitations.* As the application in quick commerce strongly defines our decision problem, we highlight the following limitations in our approach. Firstly, as we are in the realm of common LRPs that assume an average demand in a static setting, the initial solution warrants re-evaluation if demand changes drastically. This challenge holds for standard LRPs in general. While we use data from actual city maps to approximate spatial customer density, this does not cover order frequency, demand uncertainty, seasonality or demand shifts over time. Among others, our results depend on the simulated spatial MFC and customer distribution, the assumed travel distances and speed, and the applied picking and service times. Secondly, despite our usage of actual market real estate prices and publicly available processing and transportation costs, the cost parameters do not originate from a real-world application. They are subject to approximations and combinations of different sources. Thirdly, the availability of MFC locations could be very limited in an actual application as we are looking at urban neighborhoods. However, MFCs are suitable for small locales, no bigger than mid-sized shops, and with many businesses shutting down in city centers (see, e.g., Butler 2021; tagesschau.de 2023), those locations can be candidates for MFCs. Nevertheless, our approach allows the usage of a real set of available locations in the model. Fourth, quick commerce covers the central city area. Still, it could be more profitable to have a more nuanced differentiation of which areas should be prioritized and which areas might be beneficial not to service at all. Similarly, if there is already a competitor in the same city, the decision needs to be made whether we want to compete in the whole city or focus on certain neighborhoods. Lastly, new concepts for MFCs may arise in the future. Even so, as our model considers different capacity constraints, future developments in picking efficiency, including automation technologies, may be transferable to our model.

*Future Avenues of Research.* There are various extensions of our work to study additional aspects of quick commerce. As called for in (Hartl et al. 2006), there is a continuous

need to find operations research problems closely related to practical use cases. By investigating fleet management, we could extend our research into quick commerce from a strategic to an operational perspective. In our example and classical VRPs, the vehicles always return to the same depot. However, future research should consider the potential of minimizing distance travelled by allowing the vehicles to switch between depots in a dynamic setting due to the density of delivery points and depots in urban quick commerce areas. One could even reflect on allowing lateral transshipment between MFCs to redistribute inventory (Hartl and Romauch 2016) or collaboration between shippers (Gansterer and Hartl 2018; Gansterer et al. 2022), or across retailers and their warehouses (see, e.g., Mancini and Gansterer 2021). Lastly, we assume a known forecasted demand, which is static, as a starting point for this strategic problem. This has the consequence that we do not consider inventory implications, capacity shortage or surplus in our long-term strategic setting. In day-to-day operations, several inventory aspects can be investigated, such as identifying optimal inventory levels per MFC or capacity alignment between MFCs. The latter would require the application of robust approaches to the LRP.

## Appendix 1. Detailed results: city center versus outskirts area MFC network

See Tables 10 and 11.

**Table 10** Impact of MFC locations, with  $|C| = 10,000$ ,  $|L| = 10$ ,  $D = 60$ , average of 10 instances

$Q$	1		2		3		4		5	
	Base	Border	Base	Border	Base	Border	Base	Border	Base	Border
Fulfillment costs <sup>1</sup>	–	+ 13%	–	+ 8%	–	+ 5%	–	+ 3%	–	+ 2%
Setup/replen. costs <sup>2</sup>	12%	9%	16%	13%	19%	15%	21%	16%	21%	17%
Processing costs <sup>2</sup>	18%	16%	25%	23%	30%	28%	33%	32%	36%	35%
Transportation costs <sup>2</sup>	70%	75%	59%	64%	51%	57%	46%	52%	43%	48%
Lateness costs <sup>2</sup>	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%
Median arrival time <sup>3</sup>	9.48	10.15	15.24	16.65	20.56	22.03	25.30	26.90	32.68	31.49
Longest tour duration <sup>4</sup>	54.14	50.15	66.16	57.66	66.56	60.20	68.62	62.90	72.56	64.36
No. of small MFCs	7.9	7.0	7.0	8.0	7.0	8.6	7.0	8.7	6.2	8.0
No. of medium MFCs	2.0	3.0	2.0	2.0	2.0	1.4	2.0	1.0	2.0	1.1
No. of large MFCs	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

<sup>1</sup>Change of total fulfillment costs (FC) compared with base case where MFCs are located across the entire delivery area, in %

<sup>2</sup>Share of fixed setup and replenishment costs, order processing costs, transportation costs and lateness costs as % of total fulfillment costs

<sup>3</sup>Median arrival time at the last customer on a tour (incl. travel times to all customers and service times at previous customers), in min

<sup>4</sup>Longest total travel time across all tours, including return travel from last customer to the MFC, in min



**Table 11** Impact of MFC locations on quick commerce logistics, with  $|C| = 10,000, |L| = 20, D = 60$ , average of 10 instances

$Q$	1					5				
	– 10%	– 20%	– 30%	– 40%	– 50%	– 10%	– 20%	– 30%	– 40%	– 50%
Change in $c_{ini}^{fix}$ for border locations MFC										
Fulfillment costs <sup>1</sup>	– 9%	– 10%	– 11%	– 12%	– 12%	0%	– 1%	– 2%	– 3%	– 5%
Setup/replen. costs <sup>2</sup>	19%	19%	19%	18%	17%	24%	23%	22%	21%	20%
Processing costs <sup>2</sup>	20%	20%	20%	20%	20%	36%	36%	37%	37%	38%
Transportation cost <sup>2</sup>	61%	61%	61%	62%	62%	41%	41%	41%	42%	42%
Lateness costs <sup>2</sup>	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%	< 1%
Median arrival time <sup>3</sup>	6.96	6.96	6.93	6.91	6.91	29.23	29.21	29.27	29.25	29.27
Longest tour duration <sup>4</sup>	40.94	45.04	37.66	36.96	36.37	64.01	61.96	61.49	56.11	50.94
No. of small MFCs-center   border	9,007.0	9,007.0	9,007.6	9,007.8	9,007.9	5,844.9	5,845.2	5,515.5	4,916.6	4,018.0
No. of medium MFCs-center   border	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.110.0	0.000.0	0.000.0	0.000.0	0.000.0
No. of large MFCs-center   border	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0	0.000.0

<sup>1</sup>Change of total fulfillment costs (FC) compared with base case, where MFCs are located across the entire delivery area, in %<sup>2</sup>Share of fixed setup and replenishment costs, order processing costs, transportation costs and lateness costs as % of total fulfillment costs<sup>3</sup>Median arrival time at the last customer on a delivery tour (incl. travel times to and service times at previous customers), in min<sup>4</sup>Longest total travel time across all tours, including return travel from last customer to the MFC, in min

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acimovic J, Graves SC (2015) Making better fulfillment decisions on the fly in an online retail environment. *Manuf Serv Oper Manag* 17(1):34–51
- Agatz NAH, Fleischmann M, van Nunen JAEE (2008) E-fulfillment and multi-channel distribution-a review. *Eur J Oper Res* 187(2):339–356
- Aksen D, Altinkemer K (2008) A location-routing problem for the conversion to the “click-and-mortar” retailing: the static case. *Eur J Oper Res* 186(2):554–575
- Applegate DL, Bixby RE, Chvátal V, Cook WJ (2011) The traveling salesman problem. In: *The traveling salesman problem-a computational study*, Princeton university press, Princeton
- Ariker CT (2021) Do consumers punish retailers with poor working conditions during covid-19 crisis? an experimental study of q-commerce grocery retailers. *J Manag Market Logist* 8(3):140–153
- Arslan AN, Klibi W, Montreuil B (2021) Distribution network deployment for omnichannel retailing. *Eur J Oper Res* 294(3):1042–1058
- Aull B, Begley S, Chandra V, Mathur V (2021) Making online grocery a winning proposition. Last visited 2022-10-31. <https://www.mckinsey.de/industries/consumer-packaged-goods/our-insights/making-online-grocery-a-winning-proposition>
- Azadeh K, de Koster R, Roy D (2019) Robotized and automated warehouse systems: review and recent developments. *Transp Sci* 53(4):917–945
- Barreto S, Ferreira C, Paixão J, Santos BS (2007) Using clustering analysis in a capacitated location-routing problem. *Eur J Oper Res* 179(3):968–977
- Bommireddipalli RT (2022) What's next for q-commerce: the golden child of e-commerce. Last visited 2022-05-09. <https://www.forbes.com/sites/forbestechcouncil/2022/02/08/whats-next-for-q-commerce-the-golden-child-of-e-commerce/?sh=588f71394443>
- Boysen N, de Koster R, Weidinger F (2019) Warehousing in the e-commerce era: a survey. *Eur J Oper Res* 277(2):396–411
- Bretthauer KM, Mahar S, Venakataaramanan MA (2010) Inventory and distribution strategies for retail/e-tail organizations. *Comput Ind Eng* 58(1):119–132
- Buldeo Rai H, Kang S, Sakai T, Tejada C, Yuan Q, Conway A, Dablan L (2022) ‘Proximity logistics’: characterizing the development of logistics facilities in dense, mixed-use urban areas around the world. *Transp Res Part A Policy Pract* 166:41–61
- Buldeo Rai H, Marquiquoi J, Schorung M, Dablan L (2023) Dark stores in the city of light: geographical and transportation impacts of ‘Quick commerce’ in Paris. *Res Transp Econ* 100:101333
- Butler S (2021) British high street lost 11,000 shops in 2020, study shows. <https://www.theguardian.com/business/2021/mar/24/british-high-street-lost-11000-shops-in-2020-study-shows>

- Caputo AC, Pelagagge PM (2006) Management criteria of automated order picking systems in high-rotation high-volume distribution centers. *Ind Manag Data Syst* 106(9):1359–1383
- Caro F, Kök AG, Martínez-de Albéniz V (2020) The future of retail operations. *Manuf Serv Oper Manag* 22(1):47–58
- Christofides N, Eilon S (1969) An algorithm for the vehicle-dispatching problem. *J Oper Res Soc* 20:309–318
- Dablanc L, Morganti E, Arvidsson N, Woxenius J, Browne M, Saidi N (2017) The rise of on-demand ‘instant deliveries’ in European cities. *Supply Chain Forum Int J* 18(4):203–217
- Daganzo CF (1984) The distance traveled to visit  $n$  points with a maximum of  $C$  stops per vehicle: an analytic model and an application. *Transp Sci* 18(4):331–350
- Daskin MS (1997) Network and discrete location: models, algorithms and applications. *J Oper Res Soc* 48(7):763–764
- de Koster R, Le-Duc T, Roodbergen KJ (2007) Design and control of warehouse order picking: a literature review. *Eur J Oper Res* 182(2):481–501
- Dethlefs C, Ostermeier M, Hübner A (2022) Rapid fulfillment of online orders in omnichannel grocery retailing. *EURO J Transp Logist* 11:100082
- Eriksson E, Norrman A, Kembro J (2019) Contextual adaptation of omni-channel grocery retailers’ online fulfilment centres. *Int J Retail Distrib Manag* 47(12):1232–1250
- Fishman E, Cherry C (2016) E-bikes in the mainstream: reviewing a decade of research. *Transp Rev* 36(1):72–91
- Flink (2023). Flink. Last visited 2023-02-02. <https://www.goflink.com/en-DE/shop/>
- Gansterer M, Hartl RF (2018) Collaborative vehicle routing: a survey. *Eur J Oper Res* 268(1):1–12
- Gansterer M, Hartl RF, Michal T (2022) Transportation in the sharing economy. *Transp Sci* 56(3):567–570
- Gaskell TJ (1967) Bases for vehicle fleet scheduling. *J Oper Res Soc* 18(3):281–295
- Gatta V, Marcucci E, Maltese I, Iannaccone G, Fan J (2021) E-groceries: a channel choice analysis in shanghai. *Sustainability* 13(7):3625
- Getir (2023) Getir-groceries in minutes. Last visited 2023-02-02. <https://getir.uk/faq.html>
- Gund HP, Daniel J (2023) Q-commerce or e-commerce? A systematic state of the art on comparative last-mile logistics greenhouse gas emissions literature review. *Int J Ind Eng Oper Manag*
- Hartl RF, Hasle G, Janssens GK (2006) Special issue on rich vehicle routing problems. *CEJOR* 14(2):103–104
- Hartl RF, Romauch M (2016) Notes on the single route lateral transshipment problem. *J Global Optim* 65(1):57–82
- Hasani Goodarzi A, Zegordi SH (2016) A location-routing problem for cross-docking networks: a biogeography-based optimization algorithm. *Comput Ind Engineering* 102:132–146
- Held M, Karp RM (1962) A dynamic programming approach to sequencing problems. *J Soc Ind Appl Math* 10(1):196–210
- Holzapfel A, Kuhn H, Sternbeck MG (2018) Product allocation to different types of distribution center in retail logistics networks. *Eur J Oper Res* 264(3):948–966
- Hübner A, Hense J, Dethlefs C (2022) The revival of retail stores via omnichannel operations: a literature review and research framework. *Eur J Oper Res* 302(3):799–818
- Hübner A, Holzapfel A, Kuhn H, Obermair E (2019) Distribution in omnichannel grocery retailing: an analysis of concepts realized. In: Gallino S, Moreno A (eds) *Operations in an omnichannel world*, vol 8. Springer series in supply chain management. Springer Nature, Cham, pp 283–310
- Hübner A, Wollenburg J, Holzapfel A (2016) Retail logistics in the transition from multi-channel to omni-channel. *Int J Phys Distrib Logist Manag* 46(6/7):562–583
- Ishfaq R, Bajwa N (2019) Profitability of online order fulfillment in multi-channel retailing. *Eur J Oper Res* 272(3):1028–1040
- Jaghbeer Y, Hanson R, Johansson MI (2020) Automated order picking systems and the links between design and performance: a systematic literature review. *Int J Prod Res* 58(15):4489–4505
- Janjevic M, Merchán D, Winkenbach M (2021) Designing multi-tier, multi-service-level, and multi-modal last-mile distribution networks for omni-channel operations. *Eur J Oper Res* 294(3):1059–1077
- Johnson SC (1967) Hierarchical clustering schemes. *Psychometrika* 32(3):241–254
- Kale S (2021) Beware the emergency avocado: what does ultrafast delivery really cost us? Last visited 2023-02-13. <https://www.theguardian.com/business/2021/dec/09/beware-the-emergency-avocado-what-does-ultrafast-delivery-really-cost-us>

- Kang N, Shen H, Xu Y (2022) JD. com improves delivery networks by a multiperiod facility location model. *INFORMS J Appl Anal* 52(2):133–148
- Klingler R, Hübner A, Kempke T (2016) End-to-end supply chain management in grocery retailing. European Retail Institute
- Lopes RB, Barreto S, Ferreira C, Santos BS (2008) A decision-support tool for a capacitated location-routing problem. *Decis Support Syst* 46(1):366–375
- Mahar S, Modi S, Salzarulo PA (2022) Appreciating how your bread is buttered: improving online order allocation for cross-channel retailers. *Int J Prod Res* 60(15):4845–4867
- Mancini S, Gansterer M (2021) Vehicle routing with private and shared delivery locations. *Comput Oper Res* 133:105361
- Melacini M, Perotti S, Rasini M, Tappia E (2018) E-fulfilment and distribution in omni-channel retailing: a systematic literature review. *Int J Phys Distrib Logist Manag* 48(4):391–414
- Millstein MA, Bilir C, Campbell JF (2022) The effect of optimizing warehouse locations on omnichannel designs. *Eur J Oper Res* 301(2):576–590
- Millstein MA, Campbell JF (2018) Total hockey optimizes omnichannel facility locations. *INFORMS J Appl Anal* 48(4):340–356
- Min H, Current J, Schilling D (1992) The multiple depot vehicle routing problem with backhauling. *J Bus Logist* 13(1):259–288
- Nagy G, Salhi S (2007) Location-routing: issues, models and methods. *Eur J Oper Res* 177(2):649–672
- Nierynck R (2020) Quick commerce: pioneering the next generation of delivery. Last visited 2022-11-28. <https://www.deliveryhero.com/blog/quick-commerce/>
- Oezdamar L, Demir O (2012) A hierarchical clustering and routing procedure for large scale disaster relief logistics planning. *Transp Res Part E Logist Transp Rev* 48(3):591–602
- OpenStreetMap (2022) Openstreetmap. Last visited 2022-10-06. <https://www.openstreetmap.org/#map=5/51.330/10.453>
- Or I (1976) Traveling salesman-type combinatorial problems and their relation to the logistics of regional blood banking. Phd dissertation, Northwestern University, Evanston, Illinois, USA
- Partington M (2023) Speedy grocery giant gets exits Italy and Portugal—weeks after leaving Spain and France | sifted. Last visited 2023-08-27. [https://sifted.eu/articles/getir-exits-italy-and-portugal?utm\\_source=substack&utm\\_medium=email](https://sifted.eu/articles/getir-exits-italy-and-portugal?utm_source=substack&utm_medium=email)
- Perl J, Daskin MS (1985) A warehouse location-routing problem. *Transp Res Part B Methodol* 19(5):381–396
- Prodhon C, Prins C (2014) A survey of recent research on location-routing problems. *Eur J Oper Res* 238(1):1–17
- Pulido R, Muñoz JC, Gazmuri P (2015) A continuous approximation model for locating warehouses and designing physical and timely distribution strategies for home delivery. *EURO J Trans Logist* 4(4):399–419
- Rahmani Y, Ramdane Cherif-Khettaf W, Oulamara A (2016) The two-echelon multi-products location-routing problem with pickup and delivery: formulation and heuristic approaches. *Int J Prod Res* 54(4):999–1019
- Rath S, Gutjahr W (2014) A math-heuristic for the warehouse location-routing problem in disaster relief. *Comput Oper Res* 42:25–39
- Sasi G (2021) China to role model Europe's quick commerce firms. Last visited 2022-05-09. <https://appscrip.com/blog/quick-commerce-firms-should-learn-from-china/>
- Schneider M, Drexler M (2017) A survey of the standard location-routing problem. *Ann Oper Res* 259:389–414
- Simmons V, Spielvogel J, Timelin B, Tjong Pian Gi M (2022) Instant grocery: will it stay or will it go? Last visited 2022-10-31. <https://www.mckinsey.com/industries/retail/our-insights/instant-grocery-will-it-stay-or-will-it-go>
- Singh AK, Liébana-Cabanillas F (2022) An sem-neural network approach for predicting antecedents of online grocery shopping acceptance. *Int J Human Comput Interact* 1–23
- Statista (2022a) Number of SKUs of quick commerce startups 2021 | statista. Last visited 2022-10-13. <https://www.statista.com/statistics/1284960/number-skus-quick-commerce-companies/>
- Statista (2022b) Quick commerce | statista. Last visited 2022-11-25. <https://www.statista.com/study/108684/quick-commerce/>
- Statista (2023a) Commercial real estate in Europe. Last visited 2023-02-01. <https://www.statista.com/study/57836/commercial-real-estate-in-europe/>

- Statista (2023b) Quick commerce-Europa | statista marktprognose. Last visited 2023-08-27. <https://de.statista.com/outlook/dmo/online-food-delivery/grocery-delivery/quick-commerce/europa?currency=eur>
- Steinschaden J (2022) Lieferdienst flink stürzt in österreich in die pleite. Last visited 2023-08-27. <https://www.trendingtopics.eu/lieferdienst-flink-stuerzt-in-oesterreich-in-die-pleite/>
- Stone B, Davalos J (2022) The fantasy of instant delivery is imploding. Last visited 2023-02-13. <https://www.bloomberg.com/news/features/2022-10-24/gopuff-layoffs-signal-instant-delivery-s-demise>
- Tagesschau.de (2023) 9000 geschäfte dürften in diesem jahr aufgeben. <https://www.tagesschau.de/wirtschaft/verbraucher/ladensterben-corona-geschaefte-hde-handel-innenstaedte-101.html>
- Toth P, Vigo D (2014) Vehicle routing: problems, methods, and applications. Society for industrial and applied mathematics, MOS-SIAM series on optimization
- Transport Intelligence Limited (2022) The European road freight rate benchmark–Q3 2022. Last visited 2022-08-22. <https://www.ti-insight.com/european-road-freight-rate-benchmark-report/>
- Ulmer MW, Erera A, Savelsbergh M (2022) Dynamic service area sizing in urban delivery. *OR Spectr* 44(3):763–793
- Vazquez-Noguerol M, Comesaña-Benavides JA, Poler R, Prado-Prado JC (2022) An optimisation approach for the e-grocery order picking and delivery problem. *CEJOR* 30(3):961–990
- Waßmuth K, Köhler C, Agatz N, Fleischmann M (2023) Demand management for attended home delivery—a literature review. *Eur J Oper Res* 311(3):801–815
- Weber D (2021) Coronavirus pandemic permanently changes buying patterns in switzerland. Last visited 2022-10-31. <https://home.kpmg/ch/en/home/media/press-releases/2021/04/coronavirus-changes-buying-patterns-in-switzerland.html>
- Wei X, Qiu H, Wang D, Duan J, Wang Y, Cheng TCE (2020) An integrated location-routing problem with post-disaster relief distribution. *Comput Ind Eng* 147:106632
- Wollenburg J, Hübner A, Kuhn H, Trautrim A (2018) From bricks-and-mortar to bricks-and-clicks: logistics networks in omni-channel grocery retailing. *Int J Phys Distrib Logist Manag* 48(4):415–438
- Wulfraat M (2019) Micro fulfillment: passing fad or permanent market shift? Last visited 2021-06-04. [https://www.mwpl.com/MWPVL\\_International\\_-\\_Micro\\_Fulfillment\\_Centers\\_-\\_Passing\\_Fad\\_or\\_Permanent\\_Market\\_Shift.pdf](https://www.mwpl.com/MWPVL_International_-_Micro_Fulfillment_Centers_-_Passing_Fad_or_Permanent_Market_Shift.pdf)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.