Chapter 2
Network Structures Follow Network Strategies

Abstract In this chapter, we summarize and compare the structural, economic, and strategic rationales of hub-and-spoke and point-to-point network architectures; introduce the operational basics of aviation network structures; and outline the key tools needed to master complex planning and the controlling of aviation networks.

Historically, aviation networks were decentralized, regional structures. Airlines were government owned—hence the terms "national" or "flag" carrier—and served the public or political desire to provide air transport infrastructure, all based on tight regulatory rules and restrictions. After the US deregulation in 1978, highly centralized hub-and-spoke networks quickly emerged as the seemingly perfect answer to serve big markets at low costs. Rebuffing this strategy, LCCs mushroomed to attack the hub-and-spoke networks by offering direct services where the hub-and-spoke carriers only provided transfer services. Today, hub-and-spoke network architectures are playing the strengths of their network structures against LCCs, while LCCs have widely filled the vacuum of non-hubbed routes left by the hub-and-spoke carriers. Some of the largest hub-and-spoke carriers have adopted LCC hub structures, while maintaining the hub-and-spoke structure of their overall networks. In turn, LCCs are serving more transfer traffic and incorporating advanced revenue management systems to exploit the revenue potential from transfer traffic. LCC role model Southwest (WN) (Maxon 2010) is developing bank structures in Phoenix (PHX), Baltimore (BWI), Las Vegas (LAS), St. Louis (STL), Denver (DEN), and Chicago (MDW) to improve connectivity and transfer traffic. Both network structures and related business models are increasingly converging (Handelsblatt 2009). After many failed attempts, the first successful entries of LCCs on long-haul markets can be observed, while many hub-and-spoke carriers have launched LCC subsidiaries.

P. Goedeking, *Networks in Aviation*, DOI: 10.1007/978-3-642-13764-8_2,

2.1 Complying with Basic Operational Rules

Operational timing issues affect the feasible or desirable network structure of all airline networks, regardless of hub-and-spoke, LCC, or other strategic objectives behind a particular network. Therefore, we will review those operational issues—mostly rotational requirements—common to all structural network variants before studying the structural drivers of connectivity and productivity. While the differences and similarities of connectivity and production-driven network structures represent a continuum rather than distinct categories, the drivers of this continuum—connectivity versus operational efficiency—are clearly distinct in terms of strategic objective and operational levers. The following sections focus on understanding these underlying drivers rather than the resulting characteristics.

2.2 Sequencing Flights into Rotations

Passengers want to fly from A to B at particular times. Unfortunately, an airline cannot accommodate all passenger-timing preferences at the same time due to financial considerations or operational constraints. While trying to adhere to passenger demand as closely as possible, aircraft follow their specific paths through the network, mainly according to operational criteria, as do the cockpit or cabin crew. To describe such paths of aircraft, the term “rotation” is used. A rotation is a path characterized by the same airport serving as the start and final point. It may be composed of a single circle, but also may include multiple circles and other forms of paths. While the rules to build rotations for aircraft are purely technical, cockpit and cabin crews negotiate their rules in lengthy labor disputes with airline management.

Production requirements do not allow network structures to be truly distributed, but real-world networks do have centers. The crew has a home base and wishes to return to it as regularly as possible. Aircraft and maintenance facilities also are assigned to home bases. As a result, the centripetal forces of such bases create some form of centers within an airline network.

For many, the sequence of individual flights (edges) flown by a specific aircraft, along with the arrival and departure time for each destination (node), is equivalent to “the schedule.” Such “rotational plans” are typically visualized by means of Gantt-charts (see Fig. 8). Similar rotational Gantt-charts are generated for cockpit or cabin crews.

In Sect. 1.1 in Chap. 1, two theoretical types of subsequent edges were differentiated: circles (first and last nodes are the same) or paths (no more than one edge). We find paths as well as circles in airline networks. In production, circles clearly prevail, since the aircraft and crew must eventually return to their home bases.

The most common type of circle is a “ping-pong” flight. A particular aircraft flies a route in one particular direction first, and then returns to the originating point of departure. This scheme has the advantage of high operational stability:

- Aircraft frequently loop through their home base to facilitate minor maintenance work.
- Aircraft of the same type and stationed at the same base all regularly arrive or depart at short time intervals at their home base; that way, one aircraft can easily be replaced by another, making operational disruptions less likely (see Fig. 6).
- A spare aircraft at the base can be used to remedy an operational disruption.

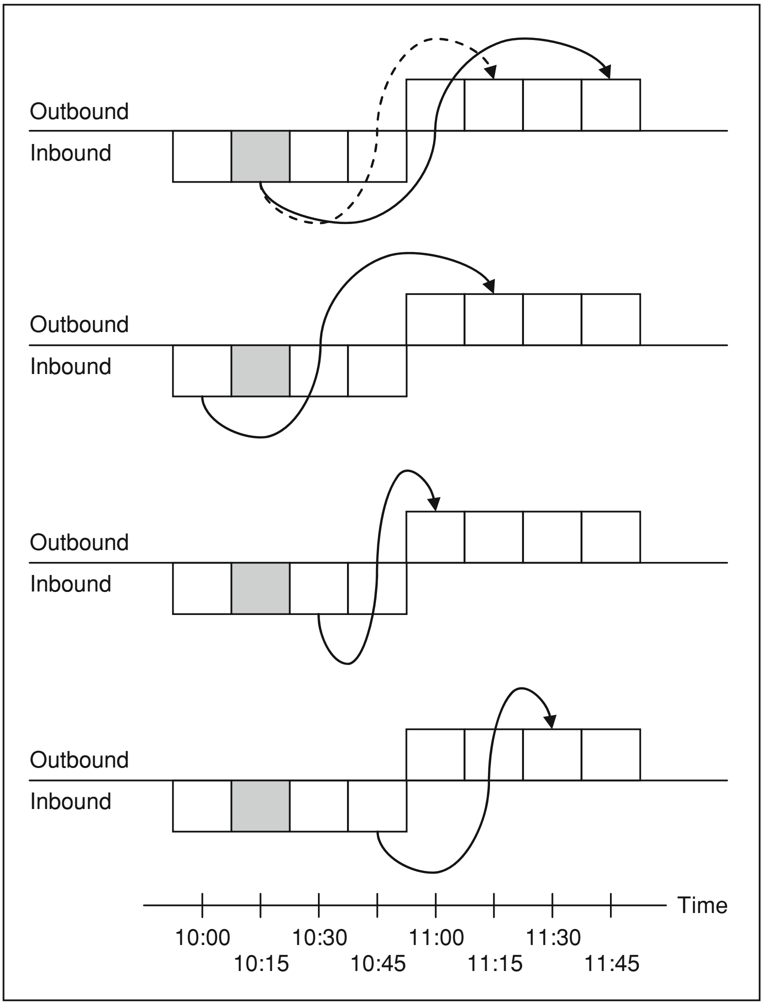

Fig. 6 Circular swapping in the event of operational disruption. The "*gray*" inbound aircraft is scheduled to depart as the second departure within the outbound wave (*dotted line*). Assuming a technical problem, this aircraft can take over only the last departure flight out of that wave (*solid line*). In this case, the sequence of rotational switches indicated in the *last three rows* can make sure all flights depart as scheduled [assuming a minimum ground time ("turnaround time," or TAT, see Sect. 2.2.1) of 30 min]

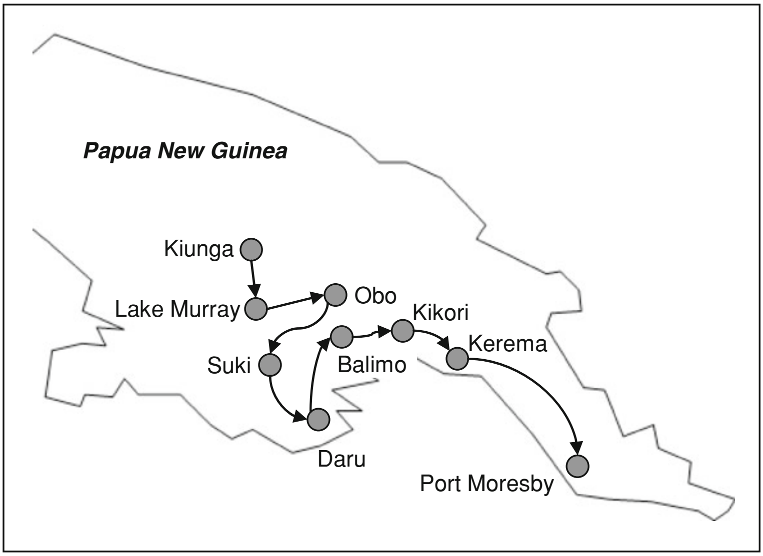

Fig. 7 A flight with eight segments. Airlines of Papua New Guinea flight number CD 359 covers a sequence of nine destinations under one single flight number

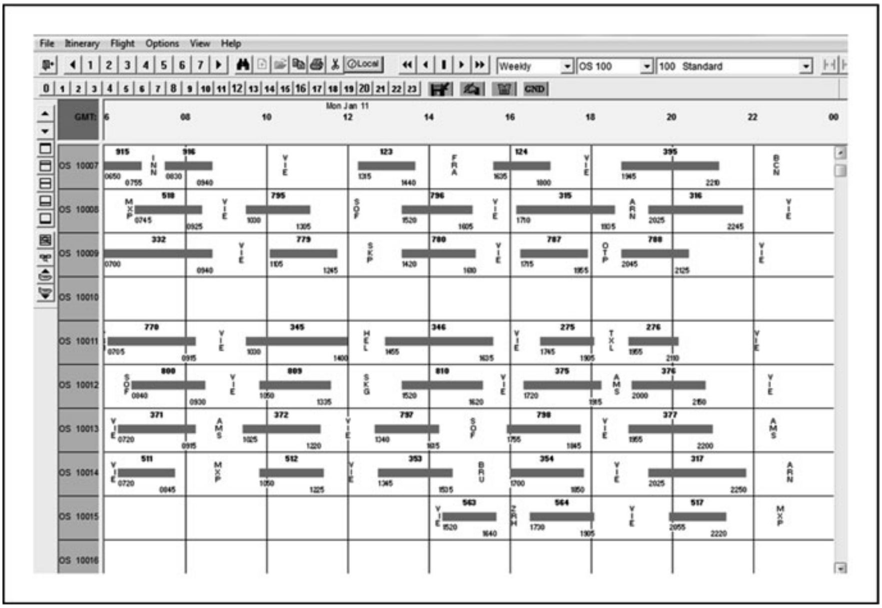

Fig. 8 A rotational map for an OS F100 aircraft (Plancor SA)

LCCs typically fly ping-pong due to their efficiency and robustness. At the other extreme, some airlines fly extended paths or rotations. A rather impressive example of such a “pearl collar” path of flights can be found on Papua New Guinea, where Airlines of Papua New Guinea (CG) serves, under one flight number (CG 359), a sequence of nine destinations (see Fig. 7), all on day seven (only).

2.2.1 Turnaround Time is Non-Productive Time

Aircraft need time after arrival before they can depart again. The time span from touching the gate (“on blocks”) until pushing back from the gate again (“off blocks”) is called the turnaround time, or TAT, of an aircraft. TATs are aircraft, airport, airline, and rotation specific. LCCs are famous for their highly efficient turnaround procedures, significantly increasing aircraft as well as terminal and apron asset utilization. TATs for a typical medium-haul aircraft such as an A320 or B737 lie in the order of 45 min, with smaller aircraft typically turning around faster, and wide-body aircraft requiring significantly more time. Some LCCs, such as Ryanair, turn a B737 around in as short a time as 20 min. Turns at an outstation are typically faster than turns at a base, where routine maintenance work is performed. TAT is different from “airport slack” (see Sect. 3.7 in Chap. 3), which is much broader and includes all time needed for taxiing, take-off/landing, and approach/climbing.

2.2.2 Building Sequences of Flights: FiFo and LiFo

Airlines typically apply two distinct methods to achieve efficient resource allocation: first-in-first-out (FiFo) and last-in-first-out (LiFo).

2.2.2.1 FiFo

Let us assume a string of inbound and outbound flights at a hypothetical airport (see Fig. 9). The task is to link (“rotate[4]”) the B737 flights in an optimal way. FiFo simply takes the first inbound flight in the morning and links it with the first available B737 outbound. Then the same procedure is repeated with the next inbound flight until all the B737 flights are exhaustively linked. Of course, once an

[4] This kind of sequencing does not necessarily lead to a complete rotation in the sense of having a starting and final airport in common. In colloquial airline terminology, however, almost any sequencing of flights is referred to as “rotating.”

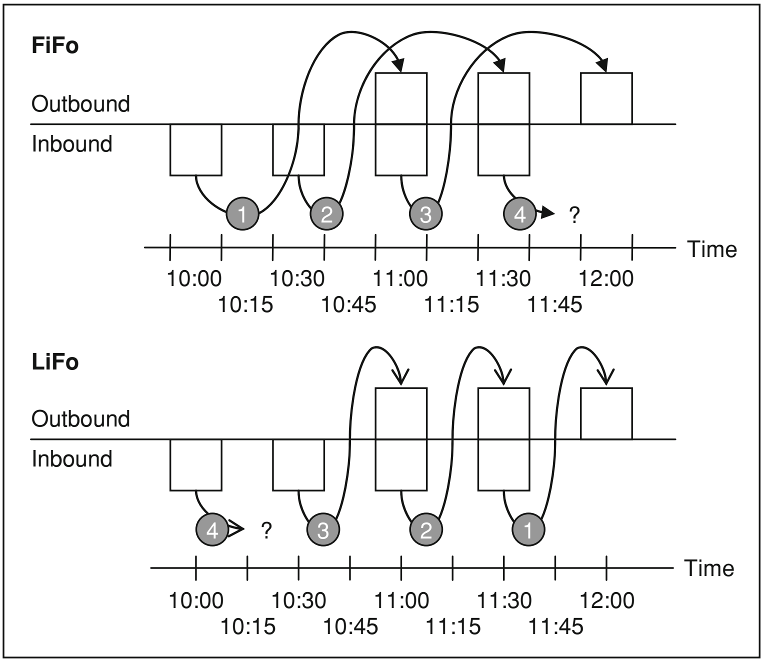

Fig. 9 FiFo and LiFo sequencing of flights

outbound flight is linked with an inbound flight the first time, it is no longer available for other links.

2.2.2.2 LiFo

Using the same example as in the FiFo case, LiFo starts with the last inbound flight (out of the B737 fleet) in the evening, linking it with the next available B737 outbound flight. The process then continues with the second last inbound, which is again connected with its respective next available (and not yet linked) outbound flight, and so on.

FiFo generates efficient, evenly packed rotational patterns; LiFo packs tighter, but sometimes opens up space. As a result, LiFo-based rotations may offer more, though limited, room to maneuver when flexibility is needed.

2.3 Hub-and-Spoke: The Answer to Deregulation?

Network airlines try to optimize connectivity, or the ability to arrive on one flight (edge) and connect to an outbound flight, thus creating a path. To do so, highly

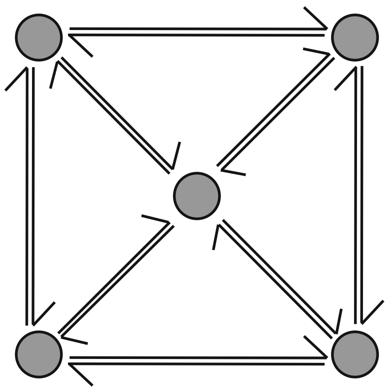

Fig. 10 In a point-to-point network, the connection of 5 nodes requires 20 flights

centric network structures are pivotal, with such centers usually referred to as "hubs." Ensuring convenient connectivity requires two key criteria: the frequency of connections offered over a particular time period (a day or a week) and the time it takes to safely connect. Given the tremendous commercial impact of hub connectivity, hubbed airlines take great care to optimize the timing of aircraft not only in terms of production efficiency, but also for network connectivity—a key aspect of product quality. The list of competitive connections within a network is a set of carefully managed paths through a few central nodes or hubs.

In the United States, airlines required the blessing of governmental "designation" before they could serve a route until 1978, when the Carter administration paved the way for full deregulation. As a result, almost all major US airlines quickly moved to replace traditional point-to-point (P2P) networks by sophisticated and connectivity-driven hub-and-spoke systems. In Asia, for instance, we still observe many highly regulated markets and governments designating national traffic routes and fares, with network structures far less advanced than their counterparts in highly competitive, deregulated environments.

Why were US airlines in the early 1980s—and European airlines soon thereafter—so eager to jump onboard hub-and-spoke systems, even though this meant asking passengers to connect on markets they traditionally flew non-stop? In other words, why did the economics of hubbing look so tempting?

Formula (2) calculates the number of routes needed in a P2P network (see Fig. 10) to connect all destinations with each other:

$$Q = n \times (n - 1) \quad (2)$$

where Q number of directional routes, n number of airports served, including hubs.

Sometimes, the formulas $Q = \frac{n \times (n-1)}{2}$ or $Q = \frac{n \times (n+1)}{2}$ can be found in the literature.

- Dividing by 2, these formulas consider routes to be bidirectional. Routes, however, are directional in both commercial and operational terms. The background of the somewhat careless use of the division by 2 is rooted in network

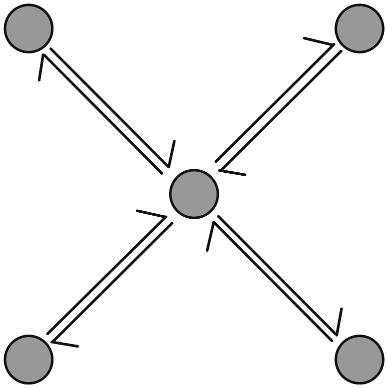

Fig. 11 In a hubbed system, five nodes can be connected, including transfer connections, by eight routes

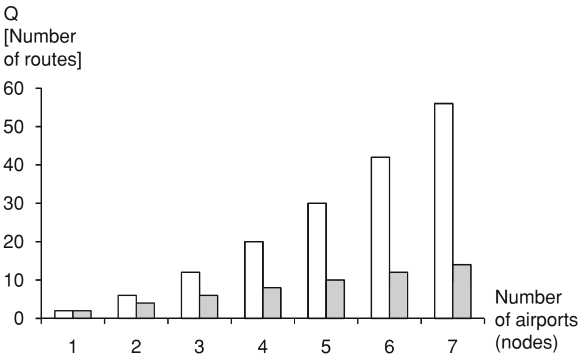

Fig. 12 Point-to-point versus hubbing. The number of routes needed in a point-to-point network [*white bars*: $Q = n(n-1)$] grows faster than the connections of a hubbed network (*gray bars*: $Q = 2(n-1)$]

topology theory where the number of edges in a network of n nodes is correctly calculated by dividing by 2, because a graphical "edge" is non-directional. Since flights or O&Ds are invariably directional, the "counting" of bidirectional edges is inappropriate in airline networks.

- The term $n \times (n+1)$ assumes that there are n airports to be connected with each other, plus a central hub (the "+1" term). While $n \times (n-1)$ assumes the hub to be included in n, the term $n \times (n+1)$ does not. Since hubs are airports as well, serving regular demand as an origin or destination, the term $n \times (n-1)$ is appropriate in the context of aviation.

The number of routes needed for a perfect hub system (see Fig. 11), where a central hub connects with all spokes, results as

$$Q = 2(n-1) \tag{3}$$

where Q = number of directional routes, and n = number of airports, including the central hub.

Figure 12 contrasts the number of routes needed to connect each airport with other airports as a function of a growing number of airports.

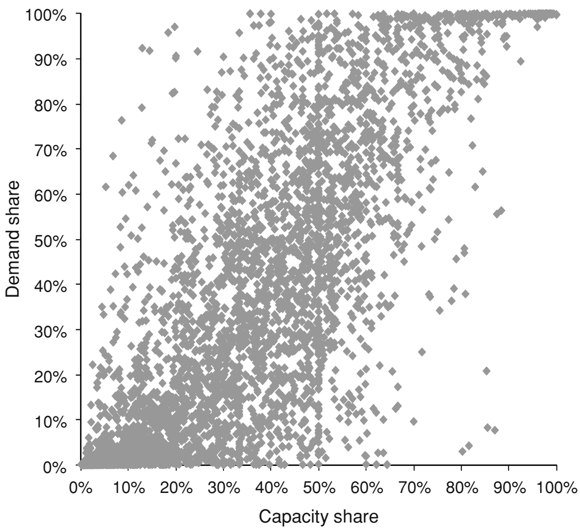

Fig. 13 The S-curve effect for 4,300 global and randomly selected O&Ds

As formulas (1) and (2) or the corresponding values in Fig. 12 make clear, the need for serviced routes develops disproportionately strong in a P2P system $(n \times (n-1))$, whereas it remains linear in a hubbed network architecture $(2 \times (n-1))$. The larger the network, the more efficiently it can be served through hubbing, assuming that demand per individual O&D is limited.

Hubs offer the chance for the dominating carrier to dominate the traffic at the respective airport. Such operational dominance can be expanded into commercial and other sorts of dominance, such as dominating brand appearance, control of corporate accounts, and price control. Often, this is referred to as the "S curve effect." This effect describes an under-proportionate share of demand as a function of a weak capacity share, whereby an over-proportionate share of capacity translates into an over-proportionate share of demand. The S curve effect must be differentiated for airports as origin, transfer hub, destination, or as per O&D or airport pair. At least at the level of O&Ds, however, the S curve effect is so weak (see Fig. 13) that it does not qualify as a strategic lever.

2.3.1 Connectivity: The Central Paradigm of Hub-and-Spoke Structures

Connectivity is defined as the ability to offer competitive connections. We will refer to competitive transfer connections as "hits." Hits must satisfy all of the following criteria:

- *Minimum connecting time* (MCT): a connection must occur within the minimum connection time, or MCT. The applicable MCTs are provided by the airports and

Arrive Station Code	Depart Station Code	Type of Connedction	Minimum Connection Time	One Wy Indicator	Connection Frequency	Arrive Carrier Code	Depart Carrier Code	Arrive Flight Number Range	Dpart Flight Number Range	Previous Station Code	Next Station Code	Effective Date	Discontinued Date
HKG	HKG	II	60	1	1234567	CX	CX		1727			20021111	29991231
HKG	HKG	II	60		1234567	CX	KA		1192			20080813	29991231
HKG	HKG	II	60		1234567	CX	KA		1196			20080813	29991231
HKG	HKG	II	60	1	1234567	CX	KA		1304		CAN	20080813	29991231
HKG	HKG	II	60		1234567	CX	KA		1306			20080813	29991231
HKG	HKG	II	60	1	1234567	CX	KA		1320		CAN	20080813	29991231
HKG	HKG	II	60	1	1234567	CX	KA		1383			20080813	29991231
HKG	HKG	II	60	1	1234567	CX	KA		1385			20080813	29991231
HKG	HKG	II	65	1	1234567	CX	LY	110	76			20090626	29991231
HKG	HKG	II	65	1	1234567	CZ	CO					20060412	29991231
HKG	HKG	II	70	1	1234567	CA	CI			PEK	TWN	20090812	29991231
HKG	HKG	II	70	1	1234567	CA	AE			TSN		20090812	29991231
HKG	HKG	II	70	1	1234567	CA	CI			TSN		20090812	29991231
HKG	HKG	II	70	1	1234567	CA	CI			TSN	TWN	20090812	29991231

Fig. 14 Example of an MCT table for HKG (September 2009, Innovata)

are updated monthly. Usually, each airport publishes a list of standard MCTs along with many exceptions. MCTs depend, for instance, on the domestic or international dimension of the connection at hand. MCT tables therefore differentiate MCTs for domestic-to-domestic (DD), domestic-to-international (DI), international-to-domestic (ID), or international-to-international (II) connections. Since connections with at least one international segment usually require customs procedures, they typically demand a longer MCT than a connection between two short-haul domestic flights. Other criteria in MCT tables relate to special flight numbers, airlines, origins, destinations, arrival or departure terminals, or validity periods. By publishing the MCT, the respective airport guarantees that all connections compliant with the rules as set out in the MCT table are feasible, for both passengers and baggage. Since many airports publish large numbers of MCT exceptions, full compliance with all MCT exceptions is indispensable when assessing connectivity. Figure 14 shows an excerpt from a typical MCT table.

Because small airports tend to have significantly shorter MCTs, they can provide faster transfer connections (see Sect. 3.7 in Chap. 3). Large airports, on the other hand, are usually much more constrained and can only offer relatively slow connections. (In Fig. 49, we present quantitative evidence for this observation.) Thus, airlines operating from smaller airports, or "hublets," can offer faster connections than those operating in larger hubs.

- *Detour*: the detour imposed by the connection must be sufficiently convenient. The detour is defined as the ratio of the total of the distances of the inbound and outbound legs, over the greater circle distance between the origin of the first leg and the destination of the second (or third in case of three segment connections).

As booking data reveal, passengers accept significantly larger detour factors for short elapsed times; while for long-haul flights, acceptable detours appear much tighter. Detour factors permit some backtracking, which is particularly relevant for long-haul connections. Reasonable detour factors for long haul are in the order of 1.2, and more generous detour factors are appropriate for short/medium haul (from 1.35 to 2.5) connections.

- *Bi-directionality*: the connection must be offered in both directions at least once per week (bi-directionality). The reason for this criterion is that it is difficult to sell a transfer connection in one direction without being able to offer a return connection—non-stop or transfer.
- *Traffic right restrictions*: as defined by IATA and as documented in the Standard Schedules Information Manual (SSIM) published by IATA (SSIM 2008), traffic right restrictions must be satisfied.

Hits must be sufficiently fast to qualify as competitive, and those that are too slow must be discarded. Hence, the question arises: how is "sufficiently fast" or "too slow" determined? There are two prevailing approaches for quantitatively assessing connectivity: One is based on a fixed time window, and the other on a flexible time window.

2.3.1.1 Fixed Maximum Connecting Time Windows

In most of the literature on this subject, the number of connections is defined by a fixed time window, referred to as "maximum connecting time," or abbreviated as MaxCT (see Fig. 15). The idea behind this concept is that meaningful connections

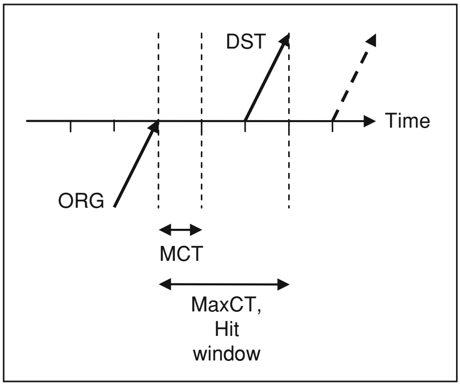

Fig. 15 Hit definition based on a fixed hit window. Connections must depart after completion of the applicable MCT, and before completion of a pre-defined time window, called maximum connecting time, or MaxCT. Additional criteria, such as detour factors, the requirement of an offer in both directions of the underlying O&D, and the absence of a faster non-stop connection, may also come into play

must happen after MCT and within the timeframe defined by MaxCT. Many different proposals exist as parameters for hit windows: Doganis and Dennis (1989) propose a standard hit window of 90 min for all types of connections. Bootsma (1997) proposes 180 min for connections between connecting continental flights, 300 min if one intercontinental flight is involved, and 720 min for connections between two intercontinental flights. Danesi (2006) suggests a differentiated set of values ranging from 90 to 180 min.

However, all approaches trying to define fixed MaxCTs have a common and significant disadvantage. A fixed hit window is of limited value for comparative purposes. A MaxCT that is competitively relevant for short/medium-haul to short/medium-haul connections in Europe is likely to be far too aggressive for many airports, even large ones, in Asia or South America. If a MaxCT of 120 min may be appropriate for Europe (medium–medium haul), at least twice as long a time window is needed to meet common and competitive transfer times of the same type at large hubs in Asia. On the other hand, if the wide-open MaxCT time windows of Asian hubs were applied to connectivity-optimized hub structures in Europe or the United States, the MaxCT would count far too many connections as being competitive.

2.3.1.2 Auto-Adaptive Hit Windows

An auto-adaptive time window takes the globally fastest possible elapsed time on a given O&D (regardless of the point of transfer), including non-stops, and at a given time period (between the time of departure and time of arrival). It takes the elapsed time of this connection as the reference, no matter how fast or slow this elapsed time may be in absolute terms (Burghouwt 2007; Burghouwt and Redoni 2009; Malighetti et al. 2008; Paleari et al. 2009). Note that for an auto-adaptive hit window, elapsed time is the reference; whereas for fixed time windows, it is the connecting time only. Any hit significantly slower (60 min in total) than the reference hit is discarded. At least conceptually, the time window to find the fastest possible connection remains open for an unlimited time period (see Fig. 16). As a result, the duration of a hit window becomes auto-adaptive toward the respective competitive environment of hits on each individual O&D worldwide. If the fastest transfer connection on a given O&D takes four hours to connect—significantly extending the resulting total elapsed time for the journey—then this elapsed time, though slow in absolute terms, is the only relevant reference for determining the competitiveness of connections on this specific O&D. For benchmarking purposes, the application of an auto-adaptive hit window or MaxCT is the only way to ensure direct comparability of the connectivity of hubs located in diverse markets. However, due to its simplicity in concept and implementation, a fixed MaxCT is preferable when focusing on structuring or evaluating schedule scenarios at a particular hub.

Connections that satisfy all of these criteria—MCT, detour, bi-directionality, granted traffic rights, and fixed or auto-adaptive hit window—are referred to as

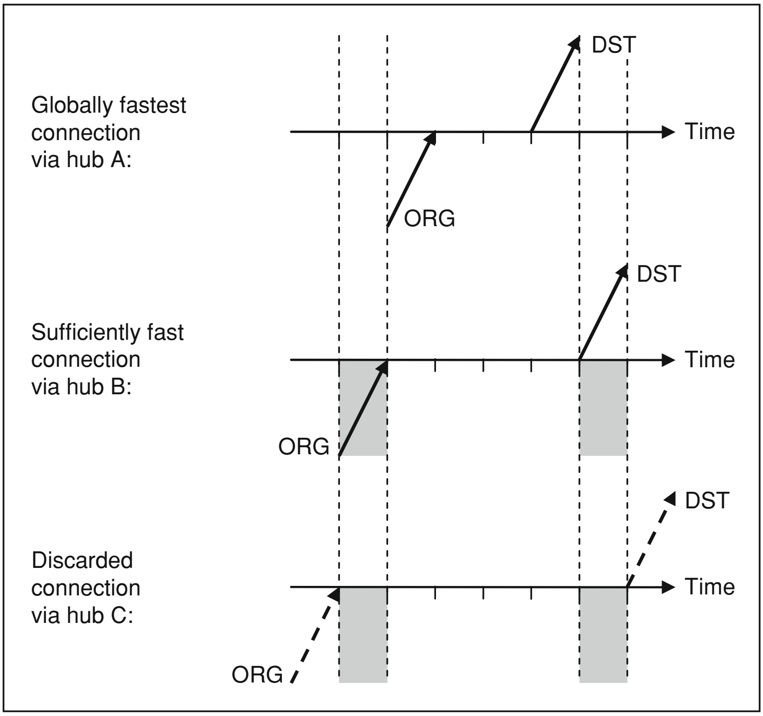

Fig. 16 Hit definition based on a flexible hit window. The duration of the hit window is defined by the duration of the globally fastest hit on the respective O&D at the time span given between scheduled time of departure (STD) of the inbound and scheduled time of arrival (STA) of the corresponding outbound flight, plus some buffer added. The auto-adaptive hit window then discards any hit slower than the timeframe set by the fastest hit and its buffer

"competitive hits" or "hits." In the framework of auto-adaptive hit windows, the consideration of detours is less relevant since an excessive detour invariably creates a slow elapsed time on a given O&D. This, in turn, makes it likely that such a connection is dominated by a less detoured connection. For the purpose of computational efficiency, however, detour factors offer advantages in the context of auto-adaptive hit windows.

2.3.2 Connectivity and Codeshares: Camouflage or Mimicry?

Frequently, airlines operating a particular flight offer the opportunity to one or more airlines to sell this flight under their own flight code (airline code plus flight number), falsely representing to the passenger that this flight is operated by the other airline. This mechanism is called "codesharing." The purpose of code-sharing is to improve sales due to two factors: (1) A flight carrying a familiar airline code sells better than a flight carrying an unfamiliar code; and (2), in GDS displays, transfer connections based on flights carrying the same airline code are displayed better than others, even if one of the two codes is based on a codeshare.

Should codeshares be considered when evaluating connectivity? If so, how? When operationally designing a hub schedule, the scheduler can only plan for the flights under the control of their airline (or online hits, where both contributing flight legs are operated by the same airline). This would suggest that for planning purposes, only operated flights may be considered. However, if the arrival or departure times of long-haul codeshares are fixed and the scheduler must plan around (de-)feeder flights, codeshares do play a central role in planning. The same applies for flights protected by antitrust-immunity (ATI). For marketing purposes, full codeshared connections are obviously beneficial. Interline connections are between flights where both legs are operated by a different airline and where no codeshare exists for these connections.

Depending on the nature of the operating or codeshared proportions, there are four distinct levels of connections:

- *Operating online*: both legs of the connection are operated by the same airline.
- *Partial codeshare*: one leg carries a codeshare, which is pivotal in establishing the respective connection.
- *Full codeshare*: both legs are codeshared only, and neither flight is operated by the airline represented by the codeshare code.
- *Interline*: the connection can only be established based on operating and/or shared codes from different airlines.

When defining hits, these levels of sharing or not sharing codes play a key role. For instance, should an operating online hit qualify as a hit if there is a faster interline connection? The applicable rules will depend on the case at hand. However, as a general rule:

- Higher levels in the connections ranking above beat lower levels.
- Faster online connections beat slower connections.

An effective way to implement such rules is by adding time penalties to the elapsed time of lower-ranking codeshares, with multiple penalties accumulating.

2.3.3 Assessing Connectivity via Connection Builders

Connection-building computer programs (CBs) exist in many variations, depending on the purpose. The Computer Reservation System (CRS) represents the most comprehensive implementation of building connections. A CRS is legally prevented from weighing connections ("biased display"), with the notable exception that certain rules apply to sort connections for display. CRSs tend to build and offer many unreasonable connections, requiring excessive detours despite the availability of more convenient connections. Typically, CRSs prioritize connections by sorting according to total elapsed travel time. To support schedule development or the competitive evaluation of schedules or schedule scenarios, CBs must apply tight definitions of competitiveness; with loose definitions, competitive connections cannot be differentiated from poor connections. A common weakness of many CBs is that they do not fully respect MCT exceptions. While often considered an "average out" shortcut, CBs can lead to severely misleading results for hubs with complex MCT rules, such as Paris CDG. Moreover, the definition of MaxCT (width of time window, parameters of fixed and/or auto-adaptive hit windows) varies widely between the various implementations of CBs. Therefore, the exact assessment of feasible hits in complex banking designs quickly becomes intricate and requires tool support (Jost 2009).

2.3.4 Evaluating Schedules with QSIs and Market Share Models

To estimate the likely market shares of a particular flight on a given O&D, one must estimate the "utility" of that flight or connection for typical passengers. One proven way is to mirror passenger decision making, based on the empirical data of real-life passengers. Let us assume a passenger wants to fly from Manchester/GB (MAN) to Istanbul/Turkey (IST). The hypothetical passenger goes to a travel agency in Manchester and asks for an option to fly from MAN to IST. The computer system in the travel agency will display a few dozen potential connections, including non-stop and transfer connections via LON, AMS, CDG, FRA, and other transfer points. What kind of evaluation criteria will the passenger apply to identify the most attractive connection? The passenger will probably compare the total travel time, the reputation of the respective airline, departure and arrival times, and the applicable airfare. Depending on the relative importance of these and other criteria, the passenger will finally opt for one particular connection. A market share model tries to emulate this rationale of decision making by quantifying all kinds of relevant quality criteria, and then applies appropriate weighting factors to each criterion. Three model families prevail:

Logit assesses the "utility" of a product or service along a series of criteria if compared against competing alternatives. "The probability of [a particular] purchase is its share of the utilities after exponentiation" (Lilien 1992; Coldren and Koppelman 2003). To apply the logit modeling to the problem of likely market shares of competing aviation itineraries, one must follow the four-step sequence of decision making used by the hypothetical passenger. The fourth step is to examine the resulting impact on likely market share. Figure 17 shows how logit models function in network analysis.

Neural networks are best understood as black boxes: they take large samples of "real" booking data to "learn" the interrelations between variables, and then apply what they have learned to predict passenger preferences for unknown or scenario markets. Neural networks yield results that in many cases are superior to logit-based models (Grosche and Rothlauf 2007). By definition, however, they do not

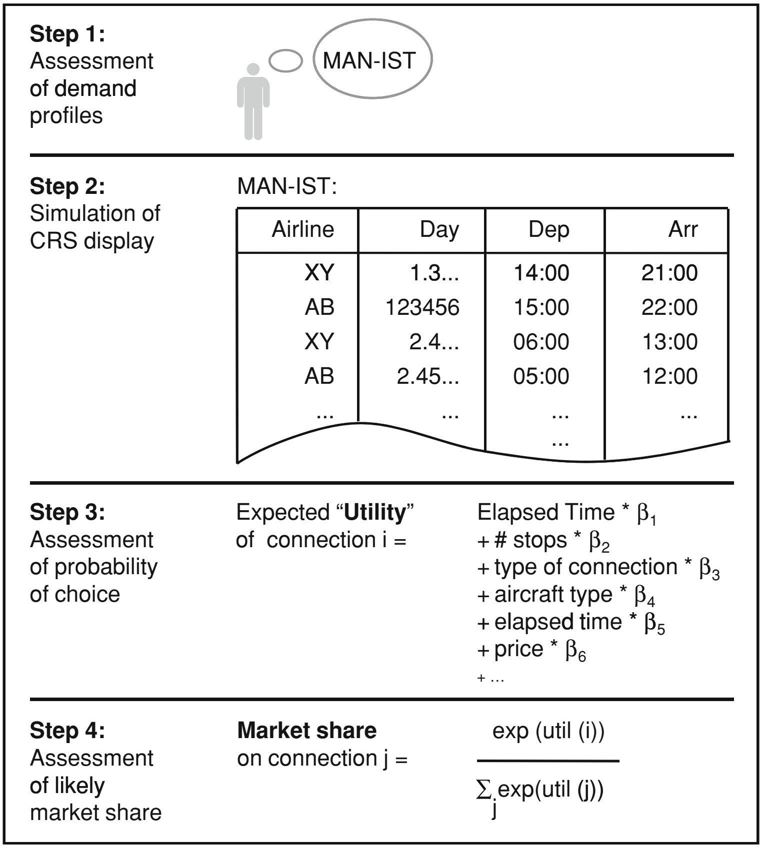

Fig. 17 The schematic rationale of logit-based market-share estimation models

provide insight into the rules that they have learned and applied. Many users of neural network based market share models shy away from accepting the outcome without understanding the reason why the model reached this result. Logit models, in contrast, offer reasonable results with maximum transparency of the computational process.

Quality of service indices follow the same rationale of mirroring passenger decision making as logit models do. These models calculate a "quality of service index" (QSI) for each connection opportunity, and take the share of the QSI score of one particular connection opportunity out of all connection opportunities on the selected O&D as a proxy for the likely resulting market share. Logit models are a variant of QSI models in that the logit model assumes a particular relationship between the factors driving the decision making.

The various QSI implementations differ in calculating what and how many criteria are used, how the relative weighting factors are determined, and how individual weighted factors are combined into one overall QSI.

2.3.5 Spill and Recapture

When planning the market performance of a network, available aircraft capacity is considered in the context of the demand on each route. With the aid of complex mathematical models, passenger demand volumes are estimated per flight segment, including behind and beyond transfer traffic. Due to operational constraints, planners frequently must assign aircraft to a route that is too small to accommodate the expected demand for that route. As a result, otherwise potential passengers are “spilled.” In large multi-hub networks, however, chances are that such “spilled” passengers will opt for a later flight or different itinerary of the same O&D served by the same airline. This effect is called “recapture” of spilled traffic. State-of-the-art planning tools take into account the recaptured traffic when estimating available demand per route.

2.4 Point-to-Point: The Answer to Hub-and-Spoke?

In highly regulated aviation markets, one finds a clear dominance of decentralized network structures with P2P services almost exclusively. An immediate effect of deregulation in the United States, or of liberalization in Europe, was that network architectures quickly shifted into significantly more centralized, or hubbed, topologies (see Fig. 18).

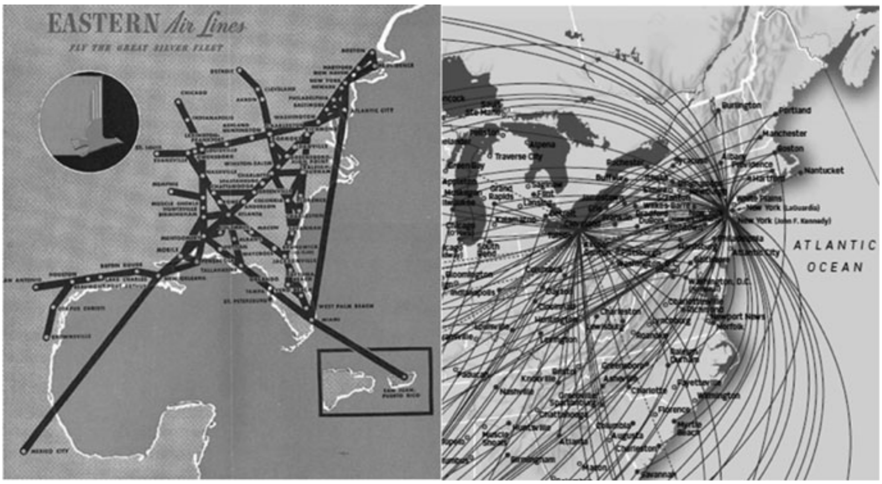

Fig. 18 Regulated versus deregulated network structures. *Left* the decentralized network structures of a prominent US pre-deregulation airline (United States Accountability Office, GAO 2006). *Right* example of a US hub-and-spoke network in 2009 (Antenna Audio Inc.)

Hub-and-spoke offered the opportunity to drastically expand the network scope—with the same fleet size inherited from former regulated times—particularly into markets too small for non-stop services, or to maintain the original network scope but with a much smaller fleet. The economics appeared simple and attractive: If an airline already serves flights from HAM to FRA and from FRA to FLR (thus covering most costs except the passenger variable cost for sales commission, peanuts, or a soft drink), passengers flying from HAM to FLR and connecting in FRA would come at marginal costs.

One caveat: when airlines observed the stunning success and strong growth of connectivity-based hub-and-spoke systems, many ordered larger aircraft to accommodate the apparently growing segment of transfer passengers. Many airlines overlooked the fact, however, that the transfer traffic only comes at marginal costs when the underlying capacity is aligned to P2P demand. When an airline opts for a larger aircraft offering more seats for transferring passengers, this capacity expansion comes at full, not marginal, costs. Many airlines had to cut back over-dimensioned aircraft when they found out about these fundamental economics. The aircraft manufacturers responded with new aircraft designs, particularly for long haul, offering relatively small capacity while maintaining high-distance reach. With these types of aircraft, airlines can afford to offer long-haul services even with limited volumes of transfer traffic, effectively keeping the full costs of transferring services at bay. At the same time, "super jumbos" like the A380 are introduced to cut down costs of high-volume, long-haul trunk routes.

In addition to the marginal cost issue, connecting passengers require many costly extra services that further complicate costs. While some of these costs are related to the spiked or "wave-like" schedule structures discussed later in Chap. 3, it is appropriate to now summarize the respective cost drivers. Here is a partial list of examples:

- Reduced asset utilization:
  - Non-optimum utilization of ground resources: Transfer-driven hub structures frequently cluster traffic into banks of high-traffic activities, typically separated by periods of comparatively low activity. All ground facilities and procedures must be tuned to maximum activity during such peak times, resulting in underutilization of resources during the periods of low activity. This applies to resources from counter positions to push-back vehicles.
  - Non-optimum utilization of aircraft and flight crew: In any such banked or "waved" system, many aircraft must wait at the preceding outstation or at the hub itself to match the timing requirements of a given inbound bank. Such idle time of the most expensive resource of an airline—aircraft plus flight crew—drains a lot of money.
- Enhanced commercial complexity:
  - Special complex algorithms in inventory or pricing systems.
  - Complex market research and competitive analysis.

- Heightened operational complexity:
  - Rapid transfer of passengers and their baggage to the departing gate also requires the airport to invest in high-performance baggage sorting and transportation devices.
  - Extra security measures and security infrastructure (note the required strict separation of Schengen- and non-Schengen passengers in Europe, requiring multimillion Euro investments from most international airports in Europe).
  - Waved patterns are vulnerable to delays: In a highly waved, spiked system, any delayed inbound flight easily creates delays or missed connections. Considering the connection-timing profile shown in Fig. 19 with its sharp rise of connections right after completion of the MCT, any delayed inbound flight would affect a few, or many, connections. Delayed inbound flights create operational frictions with regard to rotational plans for aircraft and crew, as well as for gate, runway, and air space capacities. Compare the connection timing profiles of two major US airlines at their respective prime hubs (see Fig. 19). Airline A starts building up connections as early as 25 min after arrival. The highest frequency of connections (referred to as "hits," see Sect. 2.3.1), however, is reached as late as 65 min after arrival. Airline B, in contrast, starts building up hits at 30 min and reaches its maximum only 15 min later. While the "A" pattern (solid line) is likely to offer higher scores of punctuality, the "B" structure (dotted line) is likely to appear more attractive on CRS displays.

In essence, connecting passengers can create complex costs that could severely curtail the efficiency advantages of hubbed network architectures.

The strategy and implementation of hubbed networks, while aiming at optimum coverage of as many O&Ds as possible, created its own worst enemy: LCCs. While hubbed network carriers assume that transfer O&Ds are too thin to permit direct services and cannot be served at lower costs except through hubbing, LCCs challenged both convictions. The LCC business model assumes that eliminating all the complex costs mentioned above will result in significantly lower costs, a more aggressive price point, and a boost in demand. Hence, LCCs attack hubbed

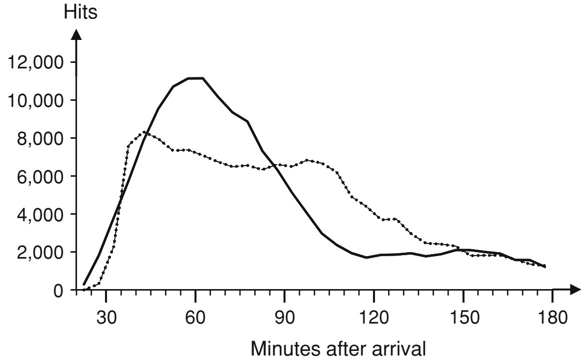

Fig. 19 Connection timing profile of two major US airlines at their respective home bases

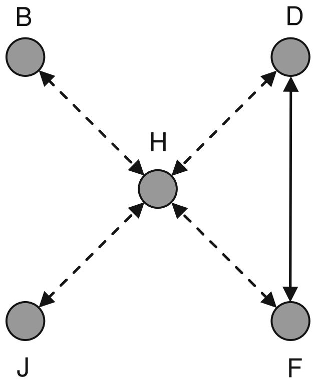

Fig. 20 LCCs (*solid line*) attack hub systems where they are most vulnerable—thin O&Ds which hub networks only serve as transfer traffic (*dotted lines*)

networks where they are most vulnerable: by serving O&Ds non-stop, which otherwise require transfers in hubbed systems (see Fig. 20). By doing so, LCCs not only grow their own business, but also simultaneously erode the demand base and the price point of their hubbed counterparts.

To remain competitive against non-stop services, hubbed carriers must offer fast connections. A slow transfer connection adds two inconveniences: the need to connect and the waste of time. The need to offer fast connections explains, at least partially, why hubbed airlines invest so heavily in sophisticated and operationally demanding high-connectivity network structures. In addition, reservation systems sort alternative connections in ascending order of total elapsed time, prioritizing faster over slower connections.

How can non-stop services compete against hubbed connections in terms of cost structure, given the enormous economies of scale inherent to hubbed systems? Frequently, successful non-stop services are offered by LCCs in direct competition with hubbed carriers.

LCCs follow different economies of scale. While network carriers leverage the massive number of city pairs they can serve at high frequency and thus generate their economies of scale, LCCs leverage the standardization of their production platform (a uniform fleet) to achieve similar or superior unit cost advantages and to permit simpler and more efficient procedures (TAT). Furthermore, LCCs exploit any other opportunity to reduce production costs. Passengers opting for the non-stop service of an LCC benefit from the convenience of a non-stop service and a more aggressive fare, but must accept the lack of differentiated service concepts on the ground and on the flight due to LCC standardization. Passengers preferring the offer of a network carrier on the same O&D may need to accept the inconvenience of a transfer, but in return have a much broader choice of connections as well as differentiated and convenient services. Increasingly, hubbed airlines and hub airports are cooperating to counter the inconvenience of connections.

Both business models, network as well as LCC carriers, are built upon economies of scale. However, both models rely on different levers to achieve such economies of scale. For the passenger, the complementary nature of both business models offers choices beyond the capability of each business model on its own. Two observations are interesting in this context:

- LCCs serve many transfer passengers and build their schedules accordingly, converging toward the business model of network carriers.
- Network carriers and hub airports, in turn, emphasize the need of high-convenience services at the point of transfer to counterbalance the intrinsic inconvenience of transfer.

2.4.1 Stuck in Between Hubs and Spokes? On Hublets

Hublets are airports with a high share of connecting traffic but built upon relatively small local demand. Thus, hublets strongly leverage—and in some cases over-leverage—limited local demand. Vienna (VIE), Austria might serve as an example of such a hublet: Austrian Airlines (OS) serves VIE airport as its homebase, serving a local demand originating in VIE in the order of 12 Mio passengers p.a., 54 destinations, and a fleet of about 27 aircraft (all 2007 figures). With this framework, OS has achieved a transfer rate of about 60% for its operations in VIE. The Vienna hublet builds upon a small catchment area, but ranks high in terms of relative connectivity if compared to the largest hubs in Europe. Without its transferring passengers, OS could not serve as many destinations and frequencies. The high rate of transfer traffic enables OS to offer a wide network of destinations and dense frequencies to local passengers in Vienna. Given the prime importance of such a broad offering of flights for the community and economy in and around Vienna, a high-performing transfer system is critical for this market. For a hub like London Heathrow (LHR), transfer traffic is a welcomed windfall profit; for hublets like VIE it is a matter of life or death. The strategic risk for hublets, however, is fragmentation. If too many and too thin O&Ds are served—and if too many small sales organizations and corporate accounts must be maintained—costs are likely to explode and exceed feasible revenues. This is particularly true if the hublet network significantly overlaps with competing hubs or hublets. The continuous decline of the PIT hublet in the north-east US over the last decade, with its limited catchment area and nearly complete overlap with surrounding hubs, depicts the vulnerability of hublet strategies.

Springer
http://www.springer.com/978-3-642-13763-1

Networks in Aviation
Strategies and Structures
Goedeking, P.
2010, XVII, 145 p., Hardcover
ISBN: 978-3-642-13763-1