Submitted to *Operations Research*

manuscript (Please, provide the mansucript number!)

Process Flexibility Revisited: The Graph Expander and Its Applications

Mabel C. Chou
NUS Business School, National University of Singapore, Singapore. Email: bizchoum@nus.edu.sg

Geoffrey A. Chua
Nanyang Business School, Nanyang Technological University, Singapore. Email: gbachua@ntu.edu.sg

Chung-Piaw Teo
NUS Business School, National University of Singapore, Singapore. Email: bizteocp@nus.edu.sg

Huan Zheng
Antai College of Economics and Management, Shanghai Jiaotong University, China. Email: zhenghuan@sjtu.edu.cn

We examine how to design a flexible process structure for a production system to match supply with demand more effectively. We argue that good flexible process structures are essentially highly connected graphs, and use the concept of graph expansion (a measure of graph connectivity) to achieve various insights into this design problem. While existing literature on process flexibility has focused on the *expected* performance of process structure, we analyze in this paper the *worst-case* performance of the flexible structure design problem under a more general setting, which encompasses a large class of objective functions. Chou et al. (2010) showed the existence of a sparse process structure which performs nearly as well as the fully flexible system on average, but the approach using random sampling yields little insights on the nature of the process structure. We show that the Ψ-expander structure, a variant of the graph expander structure (a highly connected but sparse graph) often used in communication networks, is within ϵ-optimality of the fully flexible system, *for all demand scenarios*. Furthermore, the same expander structure works uniformly well for all objective functions in our class. Based on this insight, we derive design guidelines for general non-symmetrical systems, and develop a simple and easy-to-implement heuristic to design flexible process structures. Numerical results show that this simple heuristic performs well for a variety of numerical examples previously studied in the literature, and compares favourably even with the best solutions obtained via extensive simulation and known demand distribution.

1. Introduction

Worldwide economic reforms and globalization have led to a more complex operational environment for many manufacturers. Increased reliance on make-to-order fulfillment means that manufacturers can no longer hedge against demand variability with finished goods inventory. This calls for

new production strategies that can better cope with an increasingly volatile environment. Indeed, flexibility, defined as *the ability of a system to respond or react to a change with little penalty in time, effort, or cost* (Upton, 1994), is a strategic competitive option that many manufacturers are beginning to embrace. In the automobile industry, for example, companies are moving from focused factories to flexible factories. The Ford Motor Company, for one, invested $485 million in two Canadian engine plants to retool them with a flexible system. It also plans to equip most of its 30-odd engine and transmission plants all over the world with flexible systems. Similar initiatives have also been launched in companies like GM and Nissan. Such initiatives are viewed as crucial to the survival of automakers in this increasingly competitive global environment.

The effectiveness of a flexibility strategy of this kind is highly dependent on two factors: (i) the relationship between the total invested capacity and the (random) external demand, and (ii) the design of the flexible process structure. The first issue concerns the optimal capacity to invest in, considering investment cost and demand uncertainty. The second issue revolves around how the invested capacity should be allocated among different plants, as well as what types of production capability should be configured in each plant. The focus of this paper is on the second issue.

A plant is considered more flexible if it can use its equipment and resources to produce more product types. However, how these capabilities are allocated among the plants can also affect the system's ability to handle the demand for the different products. In this setting, the focus is to design a process structure to handle as much demand as possible, or to maximize the utilization of the equipment in the plants.

The earlier studies on process flexibility basically produced two important insights. First is that if we add more flexibility to a rigid system in the right places (say, by allowing a plant to produce one more product type), a significant improvement in the system's performance can be expected. Some studies (e.g. Jordan and Graves (1995)) even provided examples showing that a very sparse partial flexibility system can be nearly as effective as a full flexibility system (where all plants can be used to produce all product types). Second, on where flexibility should be added, these studies suggested a manner of adding that creates fewer and longer chains, where a "chain" is a group of products and plants which are all connected, directly or indirectly, by product assignment decisions. Here, a long chain is preferred because it pools more plants and products and thus deals with uncertainty more effectively than a short chain. The effectiveness of the chaining strategy has been validated by many simulation studies in different areas, ranging from manpower training to

call center staffing (cf. Jordan and Graves (1995), Hopp et al. (2004), Iravani et al. (2005)).

The chaining concept by Jordan and Graves (1995) is arguably the most influential strategy used in practice to design good process structures. However, beyond the long chain, little is known about the nature of a good structure, especially for more general cases, such as when not all products and plants have the same level of mean demand and capacity. Indeed, when Jordan and Graves (1995) stated their three design principles, they also mentioned that they had no firm guidelines for adding flexibility for more general cases. That is, these design principles alone do not provide an implementable heuristic which can be used in all settings. Our paper tries to address this issue by providing a simple and implementable heuristic for the general flexibility design problem.

Our main contribution in this paper is to analyze this problem from a new perspective. In previous literature, only the average performance objectives were studied. In this paper, we analyze the performance of a sparse structure under the **worst-case** setting to ensure that our performance level can always be achieved. In addition, we generalize the model so that we can also handle objective functions such as waste minimization, as encountered in other application settings. We introduce the concept of graph expansion, which is widely used in the area of graph theory and computer science, to analyze the performance of the flexible process structure. Under a mild assumption, we show that the class of graph expanders (highly connected graphs) works extremely well for a large class of objective functions, despite the fact that it uses a far smaller number of links compared with the full flexibility system. In fact, for many classes of demand functions, we can show that the performance of 2-chain is *identical* to the performance of the fully flexible system, by analyzing its expansion properties. Finally, we use the new insights obtained from our study on graph expansion to develop new design guidelines which lead to a simple and implementable heuristic to produce a good flexible process structure.

The rest of the paper is organized as follows. In Section 2, we review the related literature on process flexibility. In Section 3, we present a general framework for the process structure design problem, encompassing the classical process flexibility model as a special case. We analyze the performance of the graph expander within our framework, when supply and demand are balanced and identical. In Section 4, we extend the result on existence of good sparse structures to non-symmetrical systems, and introduce the notion of Ψ-expander. In Section 5, we develop new design guidelines and a simple heuristic to develop good process structures for the general case when demand and supply may not be identical or balanced. In Section 6, we conduct extensive numerical

studies to illustrate the superior performance of process structures with good expansion properties, as well as to demonstrate how to implement our heuristic to construct good flexible process structures in a non-symmetrical setting. Finally, we provide some concluding remarks in Section 7.

2. Literature Review

Research on issues related to flexibility has a broad scope. Sethi and Sethi (1990) conducted an extensive survey of the applications of flexibility in different areas. They categorized 11 types of flexibility, including "machine flexibility," "product flexibility," "routing flexibility," and "resource flexibility." There is by now a vast literature in each category. Jack and Raturi (2003) studied the impact of "volume flexibility" in detail. In addition, Shi and Daniels (2003) surveyed the literature on "e-business flexibility," a new area in flexibility research. They reviewed the process flexibility literature that dealt with e-business issues and defined the concept of "e-business flexibility".

The classic work on process flexibility was conducted by Jordan and Graves (1995) based on their study of General Motors' production process. Because market conditions change quickly, customers' demand for different models is very unpredictable. The traditional "one-plant, one-model" process cannot adequately cope in this environment - demand for some models cannot be fully satisfied due to capacity limitations, whereas some plants may have spare capacity due to insufficient demand. They proposed changing the traditional focused operation to a more flexible one, where one plant can produce multiple models. In this way, the company can use the invested capacity in the plants to handle demand variations across models in a more effective manner.

The ideal design is the full flexibility system, where every plant is able to produce any product. But this is too costly, and each plant needs to have the tooling capability to produce every model. In their paper, Jordan and Graves (1995) observed (using simulations) that the partial flexibility structure, where one plant can produce only a limited number of models (suitably selected), can accrue most of the benefits offered by the full flexibility system. They further proposed a "chaining" strategy as a managerial guideline for the design of a flexibility structure.

Aksin and Karaesmen (2007) applied network theories to the study of flexible structures. The flexibility of a system is determined by the maximum network flow through customer demand to the manufacturers. They carefully studied the symmetrical flexible system and derived its submodularity property. They also derived the concavity of certain fixed process structures, as a function of the degree of each production node (the number of models each plant can handle). Hence, the returns from added flexibility into the system are diminishing.

Chou et al. (2010) demonstrated this effect more succinctly by comparing the performance of the chaining structure with the fully flexible structure for an asymptotically large system. A k-chain (denoted by $\mathcal{C}_k$) is a subgraph in an n by n bipartite graph where each supply node i is linked to demand nodes $i, i+1, \ldots, i+k-1$ (modulo n). When the demand for each product is uniformly distributed between 0 and $2C$, and each plant has a capacity of C units, they showed a surprising result that the performance of a 2-chain is already close to 89.6% of that attained by a fully flexible system when the size of the system is asymptotically large. The performance in the case of normal distribution is even more impressive. For a normal distribution $N(C, \sigma^2)$, with $C = 3\sigma$, the performance of a 2-chain goes up to an impressive level of 96%.

Many subsequent works extended the chaining strategy and partial flexibility concept and provided important observations and insights in various areas such as the supply chain (cf. Graves and Tomlin (2003), Bish et al. (2005)), flexible workforce scheduling (cf. de Farias and Van Roy (2004), Hopp et al. (2004)), and queuing (cf. Benjafaar (2002), Gurumurthi and Benjaafar (2004)). For example, Graves and Tomlin (2003) extended Jordan and Graves (1995) to obtain flexibility guidelines for multistage supply chains. On the other hand, Bish et al. (2005) cautioned that certain practices that might seem reasonable in a flexible system would lead to greater swings in production, resulting in higher operational costs, and might reduce profits.

Iravani et al. (2005) proposed a new perspective on process flexibility. They used the concept of "structural flexibility" to evaluate a system's process capability. They created an n by n "structural flexibility matrix" (SF Matrix) to study the flexibility of a cross-training CONWIP (CONstant Work-In-Process) system. They used the mean of all the elements in the SF Matrix and the dominant eigenvalue as indices of flexibility. Their research set a milestone in developing a measure for process flexibility because it allows managers to compare the performance of different process structures quickly with minimal information. Although it cannot give an absolute performance value, their paper complements ours which provides easy-to-implement methods for constructing good process flexibility structures.

Note that the system studied in our paper also belongs to a class of networks referred to as newsvendor networks that were introduced by Van Mieghem and Rudi (2002). Recently, Bassamboo et at. (2009) study the capacity and flexibility selection problem for a newsvendor network and show that the optimal configuration might go beyond chaining.

3. The Process Flexibility Problem

We use a bipartite graph to represent flexibility structures. On the left is a set $\mathcal{A}$ of n product nodes while on the right is a set $\mathcal{B}$ of m facility/plant nodes. A link connecting product node i to facility node j means that facility j has the capability to produce product i. Let $\mathcal{F} \subseteq A \times B = \{(i,j) : i \in \mathcal{A}, j \in \mathcal{B}\}$ denote the set of all such links; that is, the edge set of the bipartite graph. Hence, each flexibility configuration can be uniquely represented by a bipartite graph $\mathcal{F}$.

Let $\tilde{D}_i$ denote the demand for product i and $\tilde{\boldsymbol{D}} = (\tilde{D}_1, ..., \tilde{D}_n)$ denote the demand vector for all the products. Let $x_{i,j}$ denote the amount of demand for product i assigned to plant j and $\boldsymbol{x}$ denote the matrix of $x_{i,j}$, that is,

$$\boldsymbol{x} = (x_{i,j}), \text{ for all } i \in \mathcal{A}, j \in \mathcal{B}.$$

Let

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \triangleq \max_{\boldsymbol{x} \in \Omega_{\mathcal{F}}} \left\{ \sum_{j \in \mathcal{B}} U_j \left(\sum_{i \in \mathcal{A}} x_{i,j} \right) \right\}, \tag{1}$$

where

$$\Omega_{\mathcal{F}} = \left\{ \boldsymbol{x} : \sum_{j:(i,j) \in \mathcal{F}} x_{i,j} = \tilde{D}_i \text{ for all } i \in \mathcal{A}, \; x_{i,j} \geq 0 \text{ for all } (i,j) \in \mathcal{F}, \; x_{i,j} = 0 \text{ for all } (i,j) \notin \mathcal{F} \right\}.$$

Note that for any given flexibility structure $\mathcal{F}$ and realized demand $\tilde{\boldsymbol{D}}$, $\Omega_{\mathcal{F}}$ represents the set of all the assignments $\boldsymbol{x}$ that need to be considered in order to maximize $\sum_{j \in \mathcal{B}} U_j \left(\sum_{i \in \mathcal{A}} x_{i,j} \right)$ in (1), as explained in the following. In our model, $\sum_{i \in \mathcal{A}} x_{i,j}$ denotes the amount of demand assigned to plant j, and $U_j \left(\sum_{i \in \mathcal{A}} x_{i,j} \right)$ denotes the utility level gained by plant j from the assignment. We assume that $U_j(\cdot)$ is a non-decreasing concave utility function, but is linear in the interval $[0, C_j]$ with $U_j(0) = 0$, where C_j corresponds to the pre-configured capacity at plant j. We are thus implicitly assuming that the pre-configured capacity of the plants cannot be changed readily (as capacity investment are long term strategic decision), but can be re-deployed to meet the demand of designated products as and when needed. In addition, a non-decreasing concave utility function implies that each plant can deploy capacity beyond its pre-configured capacity from an emergency backup option with penalty to gain additional (non-negative) utility for each unit of demand it fulfills using the emergency backup option. Since this additional utility is non-negative for each plant, in order to maximize $\sum_{j \in \mathcal{B}} U_j \left(\sum_{i \in \mathcal{A}} x_{i,j} \right)$ in (1), we should always assign all demands for production because we can never be worse off by assigning an additional unit of demand to a plant. Therefore, for any given flexibility structure $\mathcal{F}$ and realized demand $\tilde{\boldsymbol{D}}$, we only consider

assignments $\boldsymbol{x}$ that assign all demands for production in order to maximize $\sum_{j\in\mathcal{B}} U_j\left(\sum_{i\in\mathcal{A}} x_{i,j}\right)$ in (1). In other words, we only consider $\boldsymbol{x}$ such that for all $i\in\mathcal{A}$, $\sum_{j:(i,j)\in\mathcal{F}} x_{i,j} = \tilde{D}_i$. Since we can not assign demand for product i to plant j unless $(i,j)\in\mathcal{F}$, the set of all the assignments $\boldsymbol{x}$ we need to consider in order to maximize $\sum_{j\in\mathcal{B}} U_j\left(\sum_{i\in\mathcal{A}} x_{i,j}\right)$ in (1) can be described as $\Omega_{\mathcal{F}}$.

Note that we assume that $U_j(x)$ is concave for utilization level beyond C_j to model the penalty associated with production beyond the pre-configured production capacity. Examples of such utility functions include:

- $U(x) = \min(x,\mu)$. Here, the plant does not gain any additional utility for production beyond μ. This models the situation when there is no emergency backup option, so that all demand beyond μ will be lost.
- $U(x) = \min(x, p+(\mu-p)x/\mu)$. Here, the plant loses a profit margin of p/μ for each unit of production beyond μ.

Note that the value $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$ depends on demand scenario $\tilde{\boldsymbol{D}}$ and process structure $\mathcal{F}$. Clearly, when $\mathcal{F}$ contains all the edges in the set $\mathcal{E} \triangleq \{(i,j) : i\in\mathcal{A}, j\in\mathcal{B}\}$, there is no restriction on which plant the demand may be assigned to, and hence the gain in utility values will be maximal. We call $\mathcal{E}$ the *fully flexible system*.

3.1. Identical and Balanced case

In this section, we assume that $|\mathcal{A}| = |\mathcal{B}| = n$ and $U(x) = U_j(x)$ for all j. It follows directly from the concavity of the objective function that

$$Z_{\mathcal{E}}(\tilde{\boldsymbol{D}}) = n\left[U\left(\frac{\sum_{i\in\mathcal{A}}\tilde{D}_i}{n}\right)\right]. \tag{2}$$

Therefore, the best strategy for $\mathcal{E}$ is to equalize the production assigned to each plant.

When $U(x) = \min(x,\mu)$, where $\mu = E(\tilde{D}_i)$, our problem reduces to the classical plant-product process design problem. A structure such as a 2-chain (denoted by $\mathcal{C}_2$) is known to work extremely well for this case.[1] In fact, asymptotically, it can be shown (cf. Chou et al. (2010)) that

$$E\left(\frac{Z_{\mathcal{C}_2}(\tilde{\boldsymbol{D}})}{n}\right) \approx 0.96 \times E\left(\frac{Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})}{n}\right) \text{ for large } n,$$

when D_i's are independent normal random variables with mean μ, standard deviation $\sigma = \mu/3$,

[1] A k-chain (denoted by $\mathcal{C}_k$) is a subgraph in an n by n bipartite graph where each supply node i is linked to demand nodes $i, i+1, \ldots, i+k-1$ (modulo n).

truncated in the range $[0, 2\mu]$. This surprising feature is desirable, because $\mathcal{C}_2$ uses a much smaller number of arcs compared to $\mathcal{E}$.

Our objective is to find a set $\mathcal{F}$ which is sparse relative to $\mathcal{E}$, that is,

$$\lim_{n\to\infty} \frac{|\mathcal{F}|}{|\mathcal{E}|} = 0,$$

but which will be able to support a production flow with a utility level as close to that of $\mathcal{E}$ as possible, *for all demand scenarios* $\tilde{\boldsymbol{D}}$. To achieve this objective, we need to find a process structure $\mathcal{F}$ which not only has very few edges, but has very high "connectivity" so that capacities can be channeled via the edges to fulfill the demands which are uncertain. Here "connectivity" means the capability of the process structure to "connect" or "link" the supply side with the demand side, thus channel the capacities to the realized demands. But since the realized demands are uncertain, in order to be able to channel the capacities to the realized demands, the edges needed may be different for different demand realizations. Thus, intuitively, to ensure higher connectivity, more edges are needed. While this intuition in a way is true, it does not tell the whole story. In particular, two graphs with the same number of edges may have different levels of connectivity depending on how the edges are assigned to connect their supply and demand nodes. For example, as pointed out in Jordan and Graves (1995), p.582, a structure with one long chain has better sales and capacity utilization performance than the structure with five short chains even though both structures have the same number of edges. The underlying reason is that the structure with one long chain has better capability in responding to unforeseen changes in demand by channeling capacities to the realized demands via the edges assigned between products and plants. Similarly, later in Section 6, Figure 1 also shows two structures which have the same number of edges but display different levels of connectivity. Therefore, it is important to assign the edges properly in order to achieve higher connectivity with the same number of edges. In this regard, it is worth noting that there is a class of highly connected graphs, called *expander*, which has received a lot of attention in the literature. Basically, expanders are graphs where every "small" subset of nodes is linked to a large neighborhood, thus allowing effective allocation of capacities to the demands. In this paper, we will use this good property of graph expanders to show how to find a set $\mathcal{F}$ which is sparse relative to $\mathcal{E}$, but which will be able to support a production flow with a utility level as close to that of $\mathcal{E}$ as possible, *for all demand scenarios* $\tilde{\boldsymbol{D}}$.

Instead of studying the average performance, we aim to find a sparse structure which performs well even under the **worst-case** demand scenario. We say that $\mathcal{F}$ is within ϵ-optimality of $\mathcal{E}$ if

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \geq (1-\epsilon) Z_{\mathcal{E}}(\tilde{\boldsymbol{D}}) \qquad \text{for all demand scenarios } \tilde{\boldsymbol{D}}.$$

We develop next a general framework for the process flexibility design problem, assuming that supply and demand are identical; that is, we assume that the demand $\tilde{D}_i$ for product i is identically distributed with mean μ, and that the capacity of each plant is pre-configured at constant μ.

The performance of the process structure depends strongly on demand variability. To the best of our knowledge, there are very few studies which take into account the impact of the variance and correlational structure of the uncertain parameters. If the variance can be arbitrarily large, then it is conceivable that a sparse process flexibility structure may be much less effective than a fully flexible structure, as demonstrated by the following example adopted from Chou et al. (2010).

EXAMPLE 1. Consider a system with n unit capacity nodes and n demand nodes, where $\tilde{D}_j = n$ with probability $1/n$ and $\tilde{D}_j = 0$ with probability $1 - 1/n$, for $j = 1, 2, \ldots, n$. Furthermore, the demands are correlated in such a way that $\sum_{j=1}^{n} \tilde{D}_j = n$ for all realizations; in other words, exactly one demand node has a value of n and all other $n-1$ demand nodes have a value of 0. Assume $U(x) = \min(x, 1)$. For any given $\tilde{\boldsymbol{D}}$, it is easy to see that in the fully flexible system, $Z_{\mathcal{E}}(\tilde{\boldsymbol{D}}) = n$. On the other hand, in any partially flexible system $\mathcal{F}$ with a degree of flexibility bounded by some fixed k (i.e., each demand node has at most k neighbors), $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$ is at most k, which is much smaller than $Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$ for a sparse process flexibility structure. □

To rule out such extreme cases, in the rest of the paper we assume that the demand satisfies the condition of *bounded variation*, defined as follows.

DEFINITION 1. $\tilde{D}_i$ has a *bounded variation* of λ if $\tilde{D}_i \leq \lambda E[\tilde{D}_i]$ for some constant λ almost surely.

It turns out that when demand has a bounded variation, we can prove that, for any given $\epsilon > 0$ and sufficiently large n, there is a process structure $\mathcal{F}$, using only a sparse number of edges, with

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \geq (1-\epsilon) Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$$

for all $\tilde{\boldsymbol{D}}$ satisfying the bounded variation condition. Intuitively, the near optimal process structure $\mathcal{F}$ identified in this paper has very few edges, but has very high connectivity with many paths[2] linking different pairs of nodes in $\mathcal{A} \cup \mathcal{B}$, thus allows us to effectively allocate capacities to the demands. To gain this intuition, we need to understand the notion of *graph connectivity* associated with every process structure.

[2] In graph theory, a path means a sequence of nodes such that from each of its nodes there is an edge to the next node in the sequence.

DEFINITION 2. Two (or more) paths are node disjoint if they have no common intermediate nodes. A structure $\mathcal{F}$ is k-connected if there are at least k node disjoint paths linking every pair of nodes in $\mathcal{A} \cup \mathcal{B}$.

There is a clear relationship between the level of connectivity and the number of edges - for higher graph connectivity, the structure needs to have more edges. A k-chain denoted by $\mathcal{C}_k$ is clearly k-connected with kn edges. However, while $\mathcal{C}_2$ is the only 2-connected graph with $2n$ edges, there are exponentially many classes of k-connected graphs with kn edges, for $k > 2$. In particular, there is a class of highly connected graphs, called the *graph expander*. The "expander" concept was first introduced by Bassalygo and Pinsker (1973) in a study of communication networks. Basically, graph expanders are graphs where every "small" subset of nodes is linked to a large neighborhood, thus allowing effective allocation of capacities to the demands. The ratio of the size of the neighborhood to the size of the subset measures the expansion capability of the graph. We define the neighborhood of a subset and the "graph expander" concept formally in the following:

DEFINITION 3. Let $\mathcal{F}$ be a bipartite graph with partite sets $\mathcal{A}$ and $\mathcal{B}$. For $S \subseteq \mathcal{A}$, the neighborhood of S in $\mathcal{F}$ is defined as

$$\Gamma_{\mathcal{F}}(S) \triangleq \{j \in \mathcal{B} : (i,j) \in \mathcal{F} \text{ for some } i \in S\}.$$

For simplicity of notation, we drop $\mathcal{F}$ and denote the neighborhood of S as $\Gamma(S)$ when there is no ambiguity about which $\mathcal{F}$ is being considered.

DEFINITION 4. Let $\mathcal{F}$ be a bipartite graph with partite sets $\mathcal{A}$ and $\mathcal{B}$. The structure $\mathcal{F}$ is an $(\alpha, \lambda, \Delta)$-expander if

- for every $v \in \mathcal{A}$, $\deg(v) \leq \Delta$, where $\deg(v)$ is the cardinality of the set $\Gamma_{\mathcal{F}}(\{v\})$, and
- for all small subsets $S \subset \mathcal{A}$ with $|S| \leq \alpha n$, we have

$$|\Gamma(S)| \geq \lambda |S|.$$

Remarks:

1. For a $n \times n$ bipartite graph which is also an $(\alpha, \lambda, \Delta)$-expander, the number of edges is at most Δn.

2. A 2-chain $\mathcal{C}_2$ is clearly a $(\frac{1}{n}, 2, 2)$-expander, since for each subset of size 1, there are at least two neighbors. Furthermore, the degree is bounded by 2. It is also a $(\frac{2}{n}, 1.5, 2)$-expander, since for every subset S of size at most 2, $|\Gamma(S)| \geq 1.5|S|$. It is easy to check that it is simultaneously a

$(\frac{k}{n}, (k+1)/k, 2)$-expander, for all $k \leq n-1$. Similarly, other graphs can be viewed as an expander with a variety of values for the triplet $(\alpha, \lambda, \Delta)$. However, we must pay attention to the values α and λ in the triplet to understand how well a graph can respond to unforeseen demand changes. In particular, α determines the largest number of nodes to be pooled together which we are interested in, and λ guarantees the minimum expansion capability of any pooled set of nodes specified by the value α. Therefore, depending on how many nodes we expect the system is able to and needs to pool due to demand uncertainties, we can set α accordingly to control the size of the pooled nodes and study the corresponding λ value to understand the expansion capability of the structure.

3. A graph expander ensures that any suitably small group of product nodes is connected to a relatively large number of plants, thus it works well in matching supply and demand as we will show in Theorem 1. Moreover, the notion that a long chain is better than a short chain can be cast in the same light: the expansion ratios for "small" subsets of product nodes in long chains are higher than those in short chains.

THEOREM 1. *Consider an $n \times n$ system, where the demand $\tilde{D}_i$ has a bounded variation of λ with mean $\mu_i = \mu$. Assume that each plant has a capacity of μ and $U(\cdot)$ is a non-decreasing concave utility function with $U(x) = Kx$ in the interval $[0, \mu]$, where K is a constant. Let $\mathcal{F}$ be an $(\alpha, \lambda, \Delta)$-expander, with $\alpha \times \lambda = 1 - \epsilon$ for some $\epsilon > 0$. Then*

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \geq \alpha\lambda n \left[U\left(\frac{\sum_{i \in \mathcal{A}} \tilde{D}_i}{n} \right) \right] = (1-\epsilon) Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$$

for **all** $\tilde{\boldsymbol{D}}$.

Proof. We start the proof with a roadmap outlining the key steps:

1. We use KKT conditions to characterize $x^*_{i,j}$, for all edge (i,j), the optimal flows between the plant and the product nodes and $U'\left(\sum_{l:l \in \mathcal{A}} x^*_{l,j}\right)$, for all plant $j \in \mathcal{B}$, the marginal utility for each plant node j. Using these characteristics, we can partition the plant nodes into groups while those plant nodes in the same group have the same marginal utility. We can then rank the groups in increasing order of its marginal utility.

2. We focus on the group with the smallest marginal utility, that is, group $\mathcal{B} \cap \mathcal{S}_1$. We note that the utility of a plant node in this group is an upper bound of the utility of any plant.

3. We focus on the group of product nodes which has $\mathcal{B} \cap \mathcal{S}_1$ as its neighbor and refer to this group as $\mathcal{T}$. That is, $\Gamma(\mathcal{T}) = \mathcal{B} \cap \mathcal{S}_1$. We consider two cases: $|\mathcal{T}| \leq \alpha n$ in case (a) and $|\mathcal{T}| \geq \alpha n$ in case (b).

4. In both cases, we use the expander property and the fact that the utility of a plant node in $\Gamma(\mathcal{T})$ is an upper bound of the utility of any plant to prove that either $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) = Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$ (in case (a)) or $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \geq \alpha\lambda n \left[U\left(\frac{\sum_{i\in\mathcal{A}} \tilde{D}_i}{n}\right)\right] = (1-\epsilon) Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$ (in case (b)).
The details of the proof are in the following.

Consider the $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$, with any given $\tilde{\boldsymbol{D}} = (\tilde{D}_1, ..., \tilde{D}_n)$. From the KKT conditions, there exists a set of lagrange multipliers $u_i^*, v_{i,j}^*$ such that the optimal solution $x_{i,j}^*$ satisfies the following conditions:

$$U'\left(\sum_{l\in\mathcal{A}} x_{l,j}^*\right) - u_i^* + v_{i,j}^* = 0 \quad \forall \quad (i,j) \in \mathcal{F} \tag{3}$$

$$\sum_{j:(i,j)\in\mathcal{F}} x_{i,j}^* = \tilde{D}_i \quad \forall \quad i \in \{1, 2, ..., n\} \tag{4}$$

$$x_{i,j}^* \times v_{i,j}^* = 0 \quad \forall \quad (i,j) \in \mathcal{F} \tag{5}$$

$$v_{i,j}^*, x_{i,j}^* \geq 0 \quad \forall \quad (i,j) \in \mathcal{F} \tag{6}$$

Let $\mathcal{S}(\tilde{\boldsymbol{D}})$ denote the support for $\boldsymbol{x}^* = (x_{i,j}^*)$; that is,

$$\mathcal{S}(\tilde{\boldsymbol{D}}) \triangleq \left\{(i,j) : x_{i,j}^* > 0\right\}.$$

Note that $\mathcal{S}(\tilde{\boldsymbol{D}}) \subseteq \mathcal{F}$.

Suppose $\mathcal{S}(\tilde{\boldsymbol{D}})$ can be written as a union of connected components $\mathcal{S}_k$, $k = 1, \ldots, h$. For each pair of nodes j and l in $\mathcal{B}$, connected to a node p in $\mathcal{A}$ in the graph induced by $\mathcal{S}_k$ (i.e., $x_{p,j}^* > 0, x_{p,l}^* > 0$), the KKT conditions (3) and (5) ensure that

$$U'\left(\sum_{i:i\in\mathcal{A}} x_{i,j}^*\right) = U'\left(\sum_{i:i\in\mathcal{A}} x_{i,l}^*\right) = u_p^*,$$

as $v_{p,l}^* = v_{p,j}^* = 0$ by (5). Since the graph $\mathcal{S}_k$ is connected,

$$U'\left(\sum_{i:i\in\mathcal{A}} x_{i,j}^*\right) = U'\left(\sum_{i:i\in\mathcal{A}} x_{i,l}^*\right)$$

for all j, l in $\mathcal{B} \cap \mathcal{S}_k$. Let β_k denote this common value. We can thus assume WLOG that $\beta_1 < \beta_2 < \ldots < \beta_h$, since we can otherwise combine components with identical β_k together. Let

$$\gamma_k \triangleq \min\{x : U'(x) = \beta_k\}. \tag{7}$$

From the definition of β_k, we can easily see that

$$\sum_{i\in\mathcal{A}} x_{i,j}^* \geq \gamma_k, \ \forall \ j \in \mathcal{B} \cap \mathcal{S}_k. \tag{8}$$

In the structure $\mathcal{F}$, we note that

$$\Gamma(\mathcal{A}\cap\mathcal{S}_1)\subseteq\mathcal{B}\cap\mathcal{S}_1. \tag{9}$$

This is because if (9) does not hold, then there exists an edge $(i,j)\in\mathcal{F}$ with $i\in\mathcal{A}\cap\mathcal{S}_1$, but $j\notin\mathcal{B}\cap\mathcal{S}_1$, which implies that either

- $j\in\mathcal{B}\cap\mathcal{S}_k$ for some $k>1$, or
- j has a flow of zero; that is, $x^*_{i,j}=0$ for all $i\in\mathcal{A}$.

But in the first case, the KKT condition (3) ensures that

$$U'\left(\sum_{l\in\mathcal{A}}x^*_{l,j}\right)-u^*_i\leq 0;$$

that is, $\beta_k\leq u^*_i$. But note that $u^*_i=\beta_1$ since $i\in\mathcal{A}\cap\mathcal{S}_1$. Therefore, $\beta_k\leq\beta_1$, which is a contradiction. In the second case, plant j is not utilized at all. Since $U(\cdot)$ is a concave function, we can always reallocate one unit of the demand for i to plant j without decreasing the value of $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$. Therefore, WLOG, we can exclude the possibility of the second case. From the above arguments, we know that (9) must hold.

Let $\mathcal{T}=\mathcal{A}\cap\mathcal{S}_1$. Since $\Gamma(\mathcal{T})\subseteq\mathcal{B}\cap\mathcal{S}_1$, and every node in $\mathcal{B}\cap\mathcal{S}_1$ is connected to some node in $\mathcal{A}\cap\mathcal{S}_1$, we have

$$\Gamma(\mathcal{T})=\mathcal{B}\cap\mathcal{S}_1. \tag{10}$$

We consider two cases - (a) and (b).

Case (a) : If $|\mathcal{T}|\leq\alpha n$, then by the expander property, $|\Gamma(\mathcal{T})|\geq\lambda|\mathcal{T}|$. Combined with (8), (10), and the bounded variation assumption, we must have

$$\lambda|\mathcal{T}|\gamma_1\leq\sum_{j\in\Gamma(\mathcal{T})}\left(\sum_{i\in\mathcal{A}}x^*_{i,j}\right)=\sum_{i\in\mathcal{T}}\tilde{D}_i\leq\lambda\mu|\mathcal{T}|.$$

Therefore, $\gamma_1\leq\mu$. Let

$$\mathcal{A}_k\triangleq\mathcal{A}\cap\mathcal{S}_k,\quad\mathcal{B}_k\triangleq\mathcal{B}\cap\mathcal{S}_k,\qquad k=1,2,...,h.$$

We consider the following three cases to show that, for all $j\in\mathcal{B}$,

$$U\left(\sum_{i\in\mathcal{A}}x^*_{i,j}\right)=K\sum_{i\in\mathcal{A}}x^*_{i,j}. \tag{11}$$

— (i): If $j \in \mathcal{B}_1$, then from (7) and the definition of β_k and $U(\cdot)$, it is easy to see that

$$U'\left(\sum_{i\in\mathcal{A}} x^*_{i,j}\right) = \beta_1 = U'(\gamma_1) = K,$$

since $\gamma_1 \le \mu$. Therefore, (11) holds.

— (ii): If $j \in \mathcal{B}_2 \cup \mathcal{B}_3 \cup ... \cup \mathcal{B}_h$, then because $U'(\cdot)$ is monotonically decreasing and $\beta_k > \beta_1$ for $k = 2, 3, ..., h$, we have $\sum_{i\in\mathcal{A}} x^*_{i,j} < \gamma_1$. Since $\gamma_1 \le \mu$, it is obvious that (11) holds for this case.

— (iii): If $j \in \mathcal{B}$, but $j \notin \mathcal{B}_1 \cup \mathcal{B}_2 \cup ... \cup \mathcal{B}_h$, then j has a flow of zero; that is, $\sum_{i\in\mathcal{A}} x^*_{i,j} = 0$. Therefore, from the definition of $U(\cdot)$, it is clear that (11) holds for this case too.

Since (11) holds for all $j \in \mathcal{B}$, from the definition of $U(\cdot)$, it is easy to see that

$$U\left(\frac{\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x^*_{i,j}}{n}\right) = \frac{K\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x^*_{i,j}}{n}.$$

Hence

$$\sum_{j\in\mathcal{B}} U\left(\sum_{i\in\mathcal{A}} x^*_{i,j}\right) = \sum_{j\in\mathcal{B}}\left(K\sum_{i\in\mathcal{A}} x^*_{i,j}\right) = K\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x^*_{i,j} = nU\left(\frac{\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x^*_{i,j}}{n}\right) = nU\left(\frac{\sum_{i\in\mathcal{A}}\tilde{D}_i}{n}\right).$$

Thus, $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) = Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$ in this case.

Case (b) : If $|\mathcal{T}| \ge \alpha n$, then $|\Gamma(\mathcal{T})|$ is at least $\alpha\lambda n = (1-\epsilon)n$. Note that

$$\sum_{i\in\mathcal{A}} x^*_{i,j} \ge \sum_{i\in\mathcal{A}} x^*_{i,k}, \text{ for all } j \in \Gamma(\mathcal{T}), k \notin \Gamma(\mathcal{T}).$$

Hence,

$$\frac{\sum_{j\in\Gamma(\mathcal{T})}\sum_{i\in\mathcal{A}} x^*_{i,j}}{|\Gamma(\mathcal{T})|} \ge \frac{\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x^*_{i,j}}{n}. \tag{12}$$

Since $U'(\sum_{i\in\mathcal{A}} x^*_{i,j})$ is a constant for all $j \in \Gamma(\mathcal{T})$, therefore, all the $\sum_{i\in\mathcal{A}} x^*_{i,j}$ with $j \in \Gamma(\mathcal{T})$ either lie in a region where the function $U(\cdot)$ is linear or lie at the same point. Combined with (12), we have

$$\sum_{j\in\Gamma(\mathcal{T})} U\left(\sum_{i\in\mathcal{A}} x^*_{i,j}\right) = |\Gamma(\mathcal{T})|U\left(\frac{\sum_{j\in\Gamma(\mathcal{T})}\sum_{i\in\mathcal{A}} x^*_{i,j}}{|\Gamma(\mathcal{T})|}\right) \ge |\Gamma(\mathcal{T})|U\left(\frac{\sum_{i\in\mathcal{A}}\tilde{D}_i}{n}\right);$$

therefore,

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \ge \alpha\lambda n\left[U\left(\frac{\sum_{i\in\mathcal{A}}\tilde{D}_i}{n}\right)\right] = (1-\epsilon)Z_{\mathcal{E}}(\tilde{\boldsymbol{D}}).$$

We have thus obtained a proof for Theorem 1. □

Note that the ϵ-optimality performance holds for all demand scenarios $\tilde{\boldsymbol{D}}$, and is thus the **worst case** performance of the expander structure, given that the demand has a bounded variation of λ.

This result is considerably stronger than the average case performance of the chaining structure. Since 2-chain $\mathcal{C}_2$ in a $n \times n$ bipartite graph is a $(\frac{n-1}{n}, \frac{n}{n-1}, 2)$-expander, we have the following immediate corollary:

COROLLARY 1. *Suppose that (i) $\tilde{D}_i$, the demand for each product i, has a bounded variation of $1 + \frac{1}{n-1}$ and has a mean $\mu_i = \mu$, $i = 1, \ldots, n$, and (ii) each of the n plants has a capacity μ. Then*

$$Z_{\mathcal{C}_2}(\tilde{\boldsymbol{D}}) = Z_{\mathcal{E}}(\tilde{\boldsymbol{D}})$$

for **all** $\tilde{\boldsymbol{D}}$.

We notice that truncated normal distribution is often used to model product demand distribution in various service and manufacturing settings. According to Corollary 1, when $\sigma = \mu/3$ and demand is truncated at one standard deviation above the mean, a 2-chain is *always* as good as the fully flexible system as long as $n \leq 4$. However, when $n \geq 4$, we note that $\mathcal{C}_2$ is a $(\frac{3}{n}, 4/3, 2)$-expander and thus its performance is $4/n$ factor of the fully flexible system in the worst case. But this implies that the worst case performance of a 2-chain is worse off compared to the fully flexible system when n increases. Therefore, for large n, we need to find a different class of graph expander structures in order to design a good process structure.

From Theorem 1, we know that an expander with α such that $\alpha\lambda = 1 - \epsilon$ has an ϵ-optimality performance. However, how many edges do we need to achieve such a performance? In other words, how big does the degree Δ need to be in order for the expander to be ϵ-optimal? We know that if Δ is as big as n, we may even have a fully flexible system. However, when n is large and Δ is much smaller than n, does there still exist such an expander with the specified α value? That is, does there *always* exist an ϵ-optimal structure with a *much smaller* number of edges than the number of edges in the fully flexible system? The answer is yes. In fact, the existence of such an expander was already proved in previous literature on graph theory, as quoted in Theorem 2.

THEOREM 2. *[Asratian et al. (1998)] For any n, $\lambda \geq 1$, and $\alpha < 1$ with $\alpha\lambda < 1$, there exists an $(\alpha, \lambda, \Delta)$-expander, for any*

$$\Delta \geq \frac{1 + \log_2 \lambda + (\lambda + 1)\log_2 e}{-\log_2(\alpha\lambda)} + \lambda + 1. \tag{13}$$

Note that the lower bound on the degree Δ is independent of n and recall that the number of edges in the expander graph is at most Δn. Hence, the number of edges in this class of graph

expanders is *linear* in n. The implication for the process flexibility problem can be stated more succinctly as follows:

In the symmetrical system, for any given demand distribution with a bounded variation of λ, we can find a corresponding α with $\alpha\lambda = 1-\epsilon$, for any given $\epsilon > 0$, such that for n sufficiently large, we can always find a process structure using at most Δn edges, where Δ is given by the right hand side of (13), such that the worst case performance of the structure is at most $1-\epsilon$ times of the fully flexible system.

We postpone the proof of Theorem 2 to the next section, where we derive a more general existence result for the non-symmetrical system using the probabilistic argument adopted from Asratian et al. (1998). While the existence of graph expanders can be established easily using the probabilistic method, the explicit construction of graph expanders proved to be much more difficult and requires a large number of sophisticated tools from number theory and graph theory. Reingold et al. (2002) used combinatorial graph product operation (zigzag product) to produce a large graph with near optimal expansion properties. We refer readers to the numerous surveys and articles for details on this subject (cf. Sarnak (2004) and the references therein).

We now consider the case when $K = 1$ in the definition of $U(x)$ and define $V(x) = x - U(x)$. Then $V(x) = 0$ for $x \le \mu$, and $V(x)$ is a non-decreasing convex function. We can define the following related problem:

$$Z'_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \triangleq \min_{\boldsymbol{x}\in\Omega_{\mathcal{F}}} \left\{ \sum_{j\in\mathcal{B}} V\left(\sum_{i\in\mathcal{A}} x_{i,j} \right) \right\},$$

where again

$$\Omega_{\mathcal{F}} = \left\{ \boldsymbol{x} : \sum_{j:(i,j)\in\mathcal{F}} x_{i,j} = \tilde{D}_i \text{ for all } i \in \mathcal{A},\ x_{i,j} \ge 0 \text{ for all } (i,j) \in \mathcal{F},\ x_{i,j} = 0 \text{ for all } (i,j) \notin \mathcal{F} \right\}.$$

In this case, our focus is on the excess demand assigned to a plant, and the penalty is increasing convex as the amount assigned moves further above μ. Interestingly, since $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$ and $Z'_{\mathcal{F}}(\tilde{\boldsymbol{D}})$ have the same feasible region, and $V(x) + U(x) = x$ for any x, we have the following result:

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) + Z'_{\mathcal{F}}(\tilde{\boldsymbol{D}}) = \sum_i \tilde{D}_i.$$

Hence, using Theorem 1, we have an analogous theorem for this class of problem:

THEOREM 3. *Let $\mathcal{F}$ be an $(\alpha, \lambda, \Delta)$-expander. When $\tilde{D}_i$ has a bounded variation of λ with mean $\mu_i = \mu$, we have*

$$Z'_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \le \alpha\lambda Z'_{\mathcal{E}}(\tilde{\boldsymbol{D}}) + (1-\alpha\lambda) \sum_i \tilde{D}_i,$$

for **all** $\tilde{\boldsymbol{D}}$. *This implies that*

$$E(Z'_{\mathcal{F}}) \leq \alpha\lambda E(Z'_{\mathcal{E}}) + (1-\alpha\lambda)n\mu.$$

4. Extension: Non-Symmetrical System

In this section, we analyze the process flexibility problem in a more general setting where demand and capacity levels are no longer identical and balanced. That is, we allow the number of product nodes and plant nodes to be different and the products to follow different demand distributions. We also allow the plants to have different capacities. To be more specific, we assume the following:

- $|\mathcal{A}| = n$ and $|\mathcal{B}| = m$, where n does not have to be equal to m.
- For all $i \in \mathcal{A}$, $E(\tilde{D}_i) = \mu_i$ and $\lambda_i^L \mu_i \leq \tilde{D}_i \leq \lambda_i^U \mu_i$ almost surely, where $0 \leq \lambda_i^L \leq 1 \leq \lambda_i^U$. We say that demand $\tilde{D}_i$ has bounded variation with λ_i^L and λ_i^U in this case.
- For all $j \in \mathcal{B}$, its pre-configured production capacity is C_j and the utility function for plant j is a concave non-decreasing function $U_j(x)$, with $U_j(x) = Kx$ for all x in $[0, C_j]$, and $U'_j(x) < K$ when $x > C_j$, to model the penalty associated with production beyond its preconfigured production capacity C_j.

Recall from (1) that our general objective is

$$Z_{\mathcal{F}}(\tilde{\boldsymbol{D}}) \triangleq \max_{\boldsymbol{x} \in \Omega_{\mathcal{F}}} \left\{ \sum_{j \in \mathcal{B}} U_j \left(\sum_{i \in \mathcal{A}} x_{i,j} \right) \right\},$$

where

$$\Omega_{\mathcal{F}} = \left\{ \boldsymbol{x} : \sum_{j:(i,j)\in\mathcal{F}} x_{i,j} = \tilde{D}_i \text{ for all } i \in \mathcal{A},\ x_{i,j} \geq 0 \text{ for all } (i,j) \in \mathcal{F},\ x_{i,j} = 0 \text{ for all } (i,j) \notin \mathcal{F} \right\}.$$

To analyze the process flexibility problem where demand and capacity levels are no longer identical and balanced, we define "Ψ-expander" as the following:

Definition 5. Given Ψ, where $0 < \Psi \leq 1$, a Ψ-expander in the process flexibility problem is a bipartite graph in $\mathcal{A} \times \mathcal{B}$ with

$$\sum_{j \in \Gamma(S)} C_j \geq \min \left\{ \sum_{i \in S} \lambda_i^U \mu_i, \Psi \sum_{j \in \mathcal{B}} C_j - \sum_{i \notin S} \lambda_i^L \mu_i \right\},$$

for all subsets $S \subseteq A$.

Given a Ψ-expander, we note that for any subset $S \subseteq A$, there are two cases:

- Case (i): $\sum_{i \in S} \lambda_i^U \mu_i \leq \Psi \sum_{j \in \mathcal{B}} C_j - \sum_{i \notin S} \lambda_i^L \mu_i$.

- Case (ii): $\sum_{i\in S}\lambda_i^U\mu_i > \Psi\sum_{j\in\mathcal{B}}C_j - \sum_{i\notin S}\lambda_i^L\mu_i$.

In Case (i), it is easy to see from Definition 5 that

$$\sum_{j\in\Gamma(S)}C_j \geq \sum_{i\in S}\lambda_i^U\mu_i,$$

and hence the plants supplying to such a subset $S\subseteq A$ have sufficient capacity to deal with the demand arising from S.

In Case (ii), we see from Definition 5 that

$$\sum_{j\in\Gamma(S)}C_j \geq \Psi\sum_{j\in\mathcal{B}}C_j - \sum_{i\notin S}\lambda_i^L\mu_i,$$

which implies that the capacity connected to such a subset S is also large enough so that at least Ψ proportion of the total capacity is utilized in the worst case.

For ease of reference, we define *small* subset as the following:

DEFINITION 6. Given a Ψ-expander, we refer to a subset $S\subseteq A$ as a *small* subset if

$$\sum_{i\in S}\lambda_i^U\mu_i \leq \Psi\sum_{j\in\mathcal{B}}C_j - \sum_{i\notin S}\lambda_i^L\mu_i.$$

For any $S\subseteq A$ that is not a *small* subset, we call it a *non-small* subset.

Combining Case (i) and (ii), we see that the definition of Ψ-expander partitions the subsets of $\mathcal{A}$ into two groups, *small* and *non-small* subsets: (i) For a *small* subset S, the plants supplying to it have sufficient capacity to deal with the demand arising from it. (ii) At the same time, the capacity connected to a *non-small* subset is also large enough so that at least Ψ proportion of the total capacity is utilized in the worst case. It is thus easy to see that a structure with $\Psi = 1$ is as good as full flexibility, and the larger Ψ is, the more flexible is a structure.

We can adapt the arguments in Section 3 to prove the following:

THEOREM 4. *Let $\mathcal{F}$ be a Ψ-expander. When $\tilde{D}_i$ has bounded variation with λ_i^L and λ_i^U for all i, then for any demand realization $\tilde{D}$, we can find a solution for $Z_{\mathcal{F}}(\tilde{D})$ such that either (a) all the plants are operating below their pre-configured capacity level (because of insufficient demand), that is, there is no performance degradation since all the demands are fulfilled in a way that generates the highest possible utility level, or (b) at least Ψ proportion of the total pre-configured capacity have been utilized.*

Proof. The detailed proof can be found in the e-companion. □

If we normalize for the demand, Theorem 4 states that a Ψ-expander has the following nice property - as long as the demand for each product falls in the range $\lambda_i^L \mu_i$ and $\lambda_i^U \mu_i$, then the process structure guarantees a utilization rate of $100 \times \Psi\%$ in the entire system!

EXAMPLE 2. Consider a setting with 5 plants and 5 products. Capacity at each plant is 100 units, whereas the demand for the 5 products are between 50 and 150, each with mean of 100. Note that we did not specify the precise structure of the demand distributions. A fully flexible system in this case contains 25 edges, whereas a 2-chain has only 10 edges. Note that the demand is always within 1.5 times of its mean. Hence the 2-chain has bounded variation with $\lambda_i^L = 0.5$, and $\lambda_i^U = 1.5$. Using Definition 5 and considering subsets $S \subseteq A$ with all possible cardinalities (from 0 to 5), we can show that the 2-chain is a 1-expander. Indeed, for any S with $|S| = 0$, $|\Gamma(S)| = 0$ and thus $\sum_{j\in\Gamma(S)} C_j = 0 \geq \sum_{i\in S} \lambda_i^U \mu_i = 0$. For any S with $|S| = 1$, $|\Gamma(S)| = 2$ and thus $\sum_{j\in\Gamma(S)} C_j = 200 \geq \sum_{i\in S} \lambda_i^U \mu_i = 150$. For any S with $|S| = 2$, $|\Gamma(S)| \geq 3$ and thus $\sum_{j\in\Gamma(S)} C_j \geq 300 \geq \sum_{i\in S} \lambda_i^U \mu_i = 300$. For any S with $|S| = 3$, $|\Gamma(S)| \geq 4$ and thus $\sum_{j\in\Gamma(S)} C_j \geq 400 \geq \Psi \sum_{j\in\mathcal{B}} C_j - \sum_{i\notin S} \lambda_i^L \mu_i = 1 * 500 - 2 * 50 = 400$. For any S with $|S| = 4$, $|\Gamma(S)| = 5$ and thus $\sum_{j\in\Gamma(S)} C_j = 500 \geq \Psi \sum_{j\in\mathcal{B}} C_j - \sum_{i\notin S} \lambda_i^L \mu_i = 1 * 500 - 1 * 50 = 450$. For any S with $|S| = 5$, $|\Gamma(S)| = 5$ and thus $\sum_{j\in\Gamma(S)} C_j = 500 \geq \Psi \sum_{j\in\mathcal{B}} C_j - \sum_{i\notin S} \lambda_i^L \mu_i = 1 * 500 - 0 * 50 = 500$. Thus the 2-chain structure in this case is a 1-expander and has the **same** performance as the fully flexible system, for all demand realizations!

In the rest of this section, we demonstrate that a sparse Ψ-expander exists for any $\Psi < 1$, provided n is sufficiently large and $E(D_i) = \mu_i = O(1)$ for each i, i.e., no single product dominates the production requirement of the system. Note that otherwise our design problem is actually easier, since considerably more capacities will be committed to support the production needs of the single product. WLOG, we can assume $\lambda_i^L = 0$ and $\lambda_i^U = \lambda$, since we can always pre-commit the capacity to produce up to the minimum level of demand for each product, thus reducing λ_i^L to 0. We assume that $|\mathcal{A}| = n$ and $|\mathcal{B}| = m$, where n does not have to be equal to m. We also assume that $\sum_j C_j \geq \sum_i \mu_i$ and that C_j, μ_i are positive integers for all i, j.

THEOREM 5. *For any $\Psi < 1$, let $\alpha = \Psi/\lambda$. Choose*

$$\Delta \geq \frac{1 + \log_2 \lambda + (\lambda + 1)\log_2 e}{-\log_2(\alpha\lambda)} + \lambda + 1. \tag{14}$$

There exists a sparse Ψ-expander structure $\mathcal{F}$, with degree $O(\Delta\mu_i) = O(1)$ at demand node i, such that

$$\sum_{j\in\Gamma_{\mathcal{F}}(S)} C_j \geq \lambda \sum_{i\in S} \mu_i,$$

for all subsets $S \subseteq A$ *with*

$$\sum_{i\in S} \mu_i \leq \alpha \sum_j C_j.$$

Proof. Consider the following probabilistic method to generate a flexibility structure: For each node i in A, pick $\Delta\mu_i$ neighbors in B randomly, with each element j sampled with probability proportional to C_j. For each set U with $\sum_{i\in U}\mu_i = z \leq \alpha\sum_j C_j$, the probability that all neighbors are contained in a set V with $\sum_{j\in V} C_j = \lambda z$ is given by

$$\prod_{i\in U} (\lambda z / \sum_j C_j)^{\Delta\mu_i} = (\lambda z / \sum_j C_j)^{z\Delta}.$$

There are at most $\binom{\sum_i \mu_i}{z}$ and $\binom{\sum_j C_j}{\lambda z}$ ways to choose U and V respectively. Hence the probability that there exist such sets U and V is at most

$$\begin{aligned} g_z &= \binom{\sum_i \mu_i}{z}\binom{\sum_j C_j}{\lambda z}(\lambda z/\sum_j C_j)^{z\Delta} \leq \binom{\sum_j C_j}{z}\binom{\sum_j C_j}{\lambda z}(\lambda z/\sum_j C_j)^{z\Delta} \\ &\leq \left(\frac{e\sum_j C_j}{z}\right)^z \left(\frac{e\sum_j C_j}{\lambda z}\right)^{\lambda z} (\lambda z/\sum_j C_j)^{z\Delta}, \end{aligned}$$

using the inequality $\binom{n}{k} \leq (ne/k)^k$. Re-arranging the terms, and using the fact that $z \leq \alpha\sum_j C_j$, we have

$$g_z \leq \left[(\sum_j C_j)^{1+\lambda-\Delta} e^{1+\lambda} \lambda^{\Delta-\lambda} z^{\Delta-\lambda-1}\right]^z \leq \left[e^{1+\lambda}\lambda(\alpha\lambda)^{\Delta-\lambda-1}\right]^z.$$

By picking Δ at least as large as the lowerbound as shown in the theorem, we can ensure that $g_z \leq (1/2)^z$. Note that $\alpha\lambda < 1$ is crucial for this to hold. Hence the probability that there exists some set U with $\sum_{i\in U}\mu_i = z \leq \alpha\sum_j C_j$, violating our condition, is at most $\sum_{z=1}^{\alpha\sum_j C_j} g_z < 1$. This proves the existence of a sparse Ψ-expander. □

5. Design Guidelines and Heuristics

We have studied the connection between worst-case performance of a process structure and graph expansion, and the existence of a sparse structure that possesses high expansion. We now use these insights to derive guidelines to design a sparse process structure given any general non-symmetrical system. To our best knowledge, this algorithmic design problem has been largely overlooked, in

part because of the technical difficulties associated with it. The only other work that attempts to tackle this issue is Jordan and Graves (1995), who provide the following three guidelines.

- Try to equalize the total capacity to which each product is directly connected.
- Try to equalize the total expected demand to which each plant is directly connected.
- Try to create a circuit visiting as many nodes as possible.

While applying these guidelines to the symmetrical case will generate regular chains (e.g. 2-chain, 3-chain, etc. depending on the budget for adding flexibility), these guidelines alone do not provide an implementable heuristic for the general non-symmetrical setting. In their 16-product, 8-plant automobile production example, Jordan and Graves (1995) added six new production links to the existing configuration, based on the above principles and by connecting products with high expected lost sales to plants with high expected excess capacity. However, to reproduce their structure is not easy, because no algorithm was provided on how to apply the guidelines. Moreover, connecting products with high expected lost sales to the most under-utilized plants requires extensive simulation. This procedure is tedious, time-consuming, and highly variable.

In this section, we first utilize the theoretical results obtained earlier to derive new design guidelines for the general non-symmetrical system. Then, based on these guidelines, we develop a simple and easy-to-implement heuristic to design flexible process structures. While our results assume bounded variation with λ_i^L and λ_i^U such that the range $[\lambda_i^L\mu_i, \lambda_i^U\mu_i]$ covers all the demand realizations, we can in practice set λ_i^L and λ_i^U more conservatively so that the range captures 80 to 90 percent of the demand. By doing this, the number of links needed will be smaller.

That said, the structural results identified in Theorem 4 are helpful if the number of *small* subsets is of manageable size. However, for a larger system, checking through all such subsets can be cumbersome. Instead, we approximate by focusing on the extremely small subsets (singletons) and the extremely large subsets (neighbors contain all but one node). Note that the definition of Ψ-expander depends on the choice of λ_i^U and λ_i^L. Ideally, we want λ_i^U to be large and λ_i^L to be small so that we can capture as much of the demand $\tilde{D}_i$ as possible within the interval $[\lambda_i^L\mu_i, \lambda_i^U\mu_i]$.

- Consider a singleton $S = \{i\}$, a *small* subset in the Ψ-expander structure. Hence, we need

$$\sum_{j\in\Gamma(S)} C_j \geq \lambda_i^U\mu_i;$$

that is, the value λ_i^U is bounded above by the following inequality:

$$\lambda_i^U \leq \frac{\sum_{j\in\Gamma(\{i\})} C_j}{\mu_i}.$$

Since we want λ_i^U to be large, we need $\frac{\sum_{j\in\Gamma(\{i\})} C_j}{\mu_i}$ to be as large as possible.

- Consider a plant node k in $\mathcal{B}$, and $T=\Gamma(\{k\})\subseteq\mathcal{A}$. Let $S=\mathcal{A}\setminus T$. S is likely to be a *non-small* subset, and hence we need

$$\sum_{j\in\Gamma(S)} C_j \geq \sum_{j\in\mathcal{B}} C_j - \sum_{i\notin S} \lambda_i^L \mu_i;$$

that is, the term $\sum_{i\notin S}\lambda_i^L\mu_i$ is bounded below by the following inequality:

$$\sum_{i\notin S}\lambda_i^L\mu_i \geq \sum_{j\in\mathcal{B}} C_j - \sum_{j\in\Gamma(S)} C_j \geq C_k.$$

If λ_i^L are identical for all $i\notin S$, then

$$\lambda_i^L \sum_{i\notin S}\mu_i \geq C_k.$$

Since we want λ_i^L to be small, we need $\frac{C_k}{\sum_{i\notin S}\mu_i}$ to be as small as possible. In other words, we need $\frac{\sum_{i\notin S}\mu_i}{C_k} = \frac{\sum_{i\in\Gamma(\{k\})}\mu_i}{C_k}$ to be as large as possible.

In summary, we provide two new design guidelines as follows.

- Make $\frac{\sum_{j\in\Gamma(\{i\})} C_j}{\mu_i}$ as large as possible for all products i.
- Make $\frac{\sum_{i\in\Gamma(\{k\})}\mu_i}{C_k}$ as large as possible for all plants k.

Next, we use the above guidelines to design our heuristic by defining the following.

Definition 7. The node-expansion ratio for $i\in\mathcal{A}$ is given by

$$\delta_i \triangleq \frac{\sum_{j\in\mathcal{B}:(i,j)\in\mathcal{F}} C_j}{E(\tilde{D}_i)}.$$

Similarly, the node-expansion ratio for $j\in\mathcal{B}$ is

$$\delta_j \triangleq \frac{\sum_{i:(i,j)\in\mathcal{F}} E(\tilde{D}_i)}{C_j}.$$

Our heuristic works by adding an edge that is not in $\mathcal{F}$ yet to increase the level of

$$\delta \triangleq \min\left\{\min_{i\in\mathcal{A}}\delta_i, \min_{j\in\mathcal{B}}\delta_j\right\}$$

as much as possible. By adding one link at a time this way, we build as much "flexibility" as possible into the system with only one additional link. By repeating this step, we can build a sparse process structure with high flexibility. Note that the heuristic, summarized in Algorithm 1, is very simple and requires minimal computational time. In fact, when adding the next link, only δ_{i^*} and δ_{j^*} need to be recomputed. Moreover, this heuristic can be further modified by examining the expansion ratios of pairs or triplets of nodes together.

ALGORITHM 1. (Expansion Heuristic: Adding a New Link)

1. Compute δ_i for each $i \in \mathcal{A}$, and δ_j for each $j \in \mathcal{B}$.
2. Set $\widehat{\mathcal{A}} := \mathcal{A}$, and $\widehat{\mathcal{B}} := \mathcal{B}$. Compute $i^* := \arg\min_{i \in \widehat{\mathcal{A}}} \delta_i$, and $j^* := \arg\min_{j \in \widehat{\mathcal{B}}} \delta_j$.
3. If $\delta_{i^*} < \delta_{j^*}$, then go to Step 4. Else, go to Step 5.
4. If $(i^*, j^*) \notin \mathcal{F}$, then $\mathcal{F} := \mathcal{F} \bigcup \{(i^*, j^*)\}$, STOP.
   Else, $\widehat{\mathcal{B}} := \widehat{\mathcal{B}} \backslash \{j^*\}$, compute $j^* := \arg\min_{j \in \widehat{\mathcal{B}}} \delta_j$, and repeat Step 4.
5. If $(i^*, j^*) \notin \mathcal{F}$, then $\mathcal{F} := \mathcal{F} \bigcup \{(i^*, j^*)\}$, STOP.
   Else, $\widehat{\mathcal{A}} := \widehat{\mathcal{A}} \backslash \{i^*\}$, compute $i^* := \arg\min_{i \in \widehat{\mathcal{A}}} \delta_i$, and repeat Step 5.

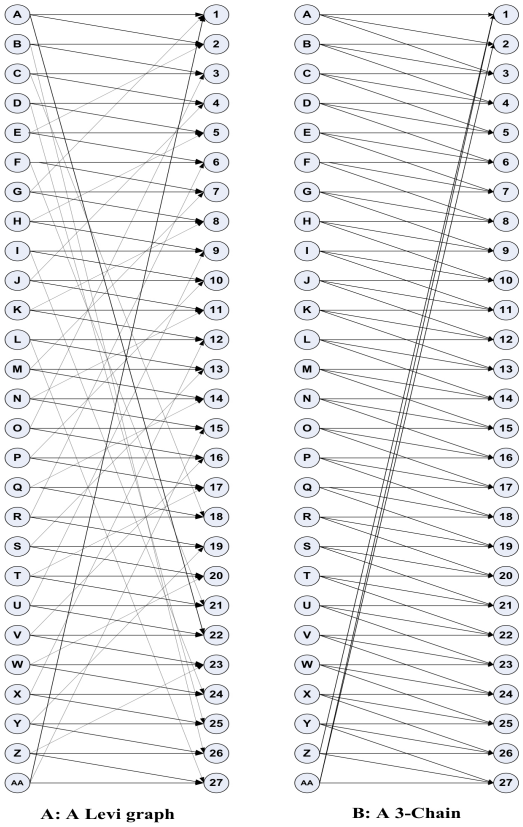
6. Numerical Studies

In this section, we conduct numerical studies to illustrate the superior performance of process structures with high expansion. We use two evaluation measures: the average performance and the worst-case performance. The former is widely used in practice and theoretical analysis, while the latter reflects a structure's robustness. We also demonstrate how to implement the heuristic developed in Section 5 on the automobile production example from Jordan and Graves (1995).

6.1. Levi graph versus the 3-chain

As shown in Figure 1, we consider a symmetrical system with 27 demand nodes and 27 plant nodes. We compare two flexibility structures, both regular graphs with degree 3 and $27 \times 3 = 81$ total links. Figure 1-B is a 3-chain while Figure 1-A is the "levi graph," well-known in graph theory for its specially selected links that ensure any two nodes share at most one common neighbor. A pair of adjacent nodes in the 3-chain, unfortunately, may have two common neighbors. Thus the levi graph has a higher expansion ratio for subsets of size not more than 2. In fact, this is also true for subsets of size not more than 3, 4, and so on until 24. According to Theorem 1, this implies that the performance of the levi graph can be guaranteed for a larger range of demand realizations than the 3-chain. We examine a symmetrical system whereby all products have identical (not necessarily independent) distributions, and all plants have the same capacity, which is equal to expected demand. Without loss of generality, we assume that mean demand is 2 for each demand node, and the capacity is also 2 for each plant. We consider 11 types of demand distributions which are a two-point distribution (demand is 1 or 4 with probabilities 2/3 and 1/3), a uniform distribution (from 0 to 4), and a variety of truncated (at 0 and 4) normal distributions with different standard

Figure 1 A levi graph and a 3-chain.



deviations (0.8, 1.2, 1.6) and correlation coefficients (0, 0.3, 0.5). Here, every 4 products fall into the same product group except for the last group which has only three products (i.e. products 1 to 4 in group 1, products 5 to 8 in group 2, etc.), and demands for products in the same group are pair-wise positively correlated according to the given correlation coefficient ρ.

For each distribution type, we generate 10,000 demand scenarios and evaluate the performances of the two structures in terms of the maximum production that the structure can support. Because the magnitude of the maximum production varies across demand scenarios, we instead keep track of performance relative to full flexibility. For example, a relative average performance of 90% means that the average maximum flow of the structure captures 90% of the average maximum flow under full flexibility. Similarly, a relative worst-case performance of 70% means that there exists one unfavorable demand scenario such that the maximum flow of the structure captures only 70% of the maximum flow under full flexibility given that same demand scenario. Table 1 summarizes the comparisons between the levi graph and the 3-chain across the different demand distributions. For the variety of demand distributions considered, we are able to observe the following patterns.

Table 1 Levi graph vs 3-chain: Summary of performance comparisons, relative to the performance of full flexibility.

Demand distributions	Average performance		Worst-Case Performance		Number of scenarios		
	Levi graph	3-chain	Levi graph	3-chain	Levi > 3-chain	Levi < 3-chain	Levi = 3-chain
$D_i = 1$ with prob. 2/3, $D_i = 4$ with prob. 1/4.	99.53%	95.85%	87.50%	77.78%	6,941	102	2,917
$D_i \sim U[0,4]$	99.78%	97.94%	90.24%	83.61%	5,977	368	3,655
$D_i \sim N(2, 0.8)$	100.00%	99.72%	98.26%	91.59%	1,868	10	8,122
$D_i \sim N(2, 1.2)$	99.97%	99.03%	96.24%	86.31%	4,023	104	5,873
$D_i \sim N(2, 1.6)$	99.92%	98.63%	94.39%	87.03%	4,917	187	4,896
$D_i \sim N(2, 0.8), \rho = 0.3$	100.00%	99.28%	95.74%	88.96%	3,190	5	6,805
$D_i \sim N(2, 1.2), \rho = 0.3$	99.95%	98.24%	93.18%	81.33%	5,118	47	4,835
$D_i \sim N(2, 1.6), \rho = 0.3$	99.91%	97.76%	92.03%	81.88%	5,850	69	4,081
$D_i \sim N(2, 0.8), \rho = 0.5$	99.99%	98.86%	95.84%	84.27%	3,866	1	6,133
$D_i \sim N(2, 1.2), \rho = 0.5$	99.94%	97.65%	92.32%	82.23%	5,690	17	4,293
$D_i \sim N(2, 1.6), \rho = 0.5$	99.87%	97.04%	88.98%	74.36%	6,271	22	3,707

The 3-chain is already very good in the average sense, netting between 95% and 99.7%, but the levi graph still manages to squeeze some improvements. However, the main advantage of the levi graph begins to show in the worst-case performance comparisons, where the gap between the levi graph and the 3-chain is quite significant, ranging from 7 to 14 percentage points. In fact, the levi graph is so good that our data show that it is as good as full flexibility in 86% of the scenarios under the 2-point distribution, 83% under uniform distribution, and 91-99% under the family of normal distributions. For the 3-chain, these numbers are 28%, 39%, and 37-81%, respectively.

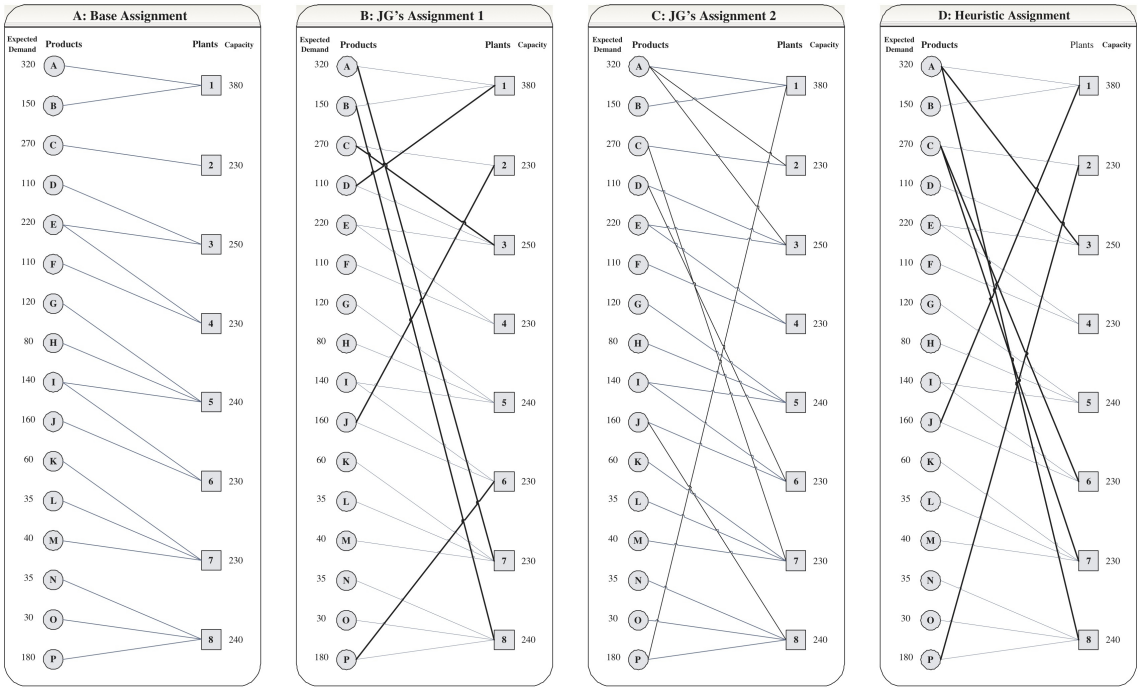
The penultimate column of Table 1 shows that the 3-chain seldom outperforms the levi graph. More importantly, these instances do not occur in the worst case. Furthermore, the worst-case performance of the 3-chain is more sensitive than the levi graph to changes in demand variance and correlation. In summary, we observe that the levi graph has better and more robust performances compared to the 3-chain, and we attribute this to the levi graph's higher expansion ratio.

6.2. Jordan and Graves' Automobile Production Example

The objective of this section is to demonstrate the implementability of the heuristic we developed in Section 5. To this end, we revisit the 16-product, 8-plant automobile production example in Jordan and Graves (1995). Figure 2-A shows the set of products with their respective expected demands, the set of plants with their respective capacities, as well as the base assignment which represents

the existing configuration of production capabilities. Suppose we have a budget to add six new

Figure 2 The structures studied in Jordan and Graves (1995) vs. the structure generated by our heuristic.



links. We can then employ our heuristic to add the following links: (C,7), (A,8), (P,2), (J,1), (C,6), (A,3), presented in the order by which they are to be added. For example, if the budget is reduced to only five new links, then (A,3) will be excluded. Figure 2-D shows the resulting structure. At this point, we note the ease of use and exactness of our heuristic since it only uses information on the capacities and expected demands and does not require any simulation of demand realizations.

We conduct a numerical study on 11 types of demand distributions and compare with the two structures proposed in Jordan and Graves (1995). These structures, shown in Figure 2-B and C, were constructed by connecting products with most lost sales to the most under-utilized plants based on extensive simulation for estimating expected lost sales and expected utilized capacities. We call these structures JG1 and JG2. We consider 5 types of independent demand distributions and 6 types of correlated distributions. The independent distributions are a two-point distribution, a uniform distribution, and a family of truncated normal distributions with different coefficients of variation (CV = 0.4, 0.6, 0.8). For the correlated distributions, we follow Jordan and Graves (1995) and divide the product nodes into three groups: Group 1 from Nodes A to F, Group 2 from Nodes G to M, and Group 3 from Nodes N to P. Grouped products are pair-wise positively correlated, but

Table 2 JG Example: Summary of performance comparisons, relative to the performance of full flexibility.

Demand distributions	Average performance					Worst-Case Performance				
	Heuristic	JG1	JG2	Heuristic $\geq$		Heuristic	JG1	JG2	Heuristic $\geq$	
				JG1	JG2				JG1	JG2
$D_i = 0$ with prob. 1/2, $D_i = 2\mu_i$ with prob. 1/2.	89.85%	88.28%	89.36%	√	√	60.78%	58.71%	59.20%	√	√
$D_i \sim U[0, 2\mu_i]$	97.30%	96.85%	97.31%	√		79.37%	78.05%	78.04%	√	√
$D_i \sim N(\mu_i, 0.4\mu_i)$	99.17%	99.17%	99.27%	√		83.05%	84.26%	87.90%		
$D_i \sim N(\mu_i, 0.6\mu_i)$	98.38%	98.24%	98.49%	√		81.70%	79.92%	83.32%	√	
$D_i \sim N(\mu_i, 0.8\mu_i)$	97.95%	97.65%	98.01%	√		80.06%	76.44%	82.82%	√	
$D_i \sim N(\mu_i, 0.4\mu_i), \rho = 0.3$	99.35%	99.32%	99.39%	√		78.27%	82.42%	82.34%		
$D_i \sim N(\mu_i, 0.6\mu_i), \rho = 0.3$	98.73%	98.64%	98.75%	√		79.21%	81.33%	79.48%		
$D_i \sim N(\mu_i, 0.8\mu_i), \rho = 0.3$	98.45%	98.26%	98.42%	√	√	78.01%	78.68%	80.14%		
$D_i \sim N(\mu_i, 0.4\mu_i), \rho = 0.5$	99.37%	99.34%	99.37%	√	√	83.01%	82.88%	82.51%	√	√
$D_i \sim N(\mu_i, 0.6\mu_i), \rho = 0.5$	98.89%	98.70%	98.79%	√	√	80.69%	78.28%	80.91%	√	
$D_i \sim N(\mu_i, 0.8\mu_i), \rho = 0.5$	98.52%	98.31%	98.43%	√	√	79.67%	78.00%	77.72%	√	√

independent of products in other groups. We consider two levels of demand correlation ($\rho = 0.3, 0.5$) for the normal distribution with three levels of variation (CV = 0.4, 0.6, 0.8).

For each distribution type, we generate 10,000 demand scenarios and evaluate the performance of all three structures in terms of the maximum production that can be supported by the structure. As in Section 6.1, we use relative average and worst-case performances to reflect how close the performances are to full flexibility. The results are shown in Table 2. For all distribution types, our heuristic structure is at least as good as the JG1 structure in the average sense. It also manages to be so against JG2 for the high-correlation types and the high-variance-with-correlation types. However, in the worst-case sense, there is a toss up between the heuristic structure and the JG structures, where the heuristic structure seems to perform better under high correlation and under high variance[3]. Moreover, JG2 appears to be slightly more superior to JG1. In general (in both average and worst-case sense), the heuristic structure performs relatively better under high correlation. However, these findings must be taken with a grain of salt. Closer scrutiny reveals that the differences in performance are only a couple of percentage points, quite often only a fraction of a percentage point. That said, given the amount of numerical simulations conducted, we are confident that the heuristic structure performs just as well as the two JG structures.

[3] Two-point and uniform distributions have higher variances than normal distribution

Table 3 JG Example: Summary of node-expansion and pair-expansion ratios.

	Node-expansion ratios			Pair-expansion ratios		
	Heuristic	JG1	JG2	Heuristic	JG1	JG2
Lowest ratio	1.4167	1.4167	1.4167	1.2000	1.1721	1.2000
2nd lowest ratio	1.4348	1.4348	1.4348	1.2717	1.2000	1.2979
3rd lowest ratio	1.6579	1.5263	1.6875	1.4113	1.3306	1.3936

So how does one explain the comparable performance among the three structures? Should the heuristic structure not perform better because it was designed to have good expansion properties? To address these questions, we compute the node-expansion ratios and obtain the lowest such ratio among all plants and products, for each of the three structures. Table 3 summarizes these lowest ratios together with the 2nd and 3rd lowest ratios, as well as the corresponding ratios for pairs of products and pairs of nodes. Interestingly, we find that the lowest node-expansion ratios for all three structures are equal at 1.4167. This explains why these structures perform almost equally well in our numerical study. For pairs of products or plants, the lowest ratio for JG1 turns out to be lower than those of JG2 and the heuristic structure, but not by much. This is consistent with the numerical observations that JG1 performs slightly worse than the heuristic structure and JG2. In summary, we argue that the main reason why all three structures perform equally well is because they all have good expansion properties, which confirms the theoretical results in this paper.

While our heuristic structure performs about just as well as the JG structures, it is important to note that our structure was constructed using a computationally efficient and exact method. It only uses information on mean demand to design the process structure. In other words, our approach is independent of the distributional information and correlational structures of the demand process. In contrast, the method proposed by Jordan and Graves conducts simulation using the actual demand distribution in order to add new links to the structure. This means that more demand information is required, the method is quite computationally expensive, and the structure generated is highly variable (e.g. one does not know whether he will obtain JG1, JG2, or another structure). However, in terms of performance, a simple, easy-to-implement method like our heuristic can deliver just as well, primarily because our method exploits the system's expansion property.

7. Conclusions

In this paper, we examine how to design a flexible process structure for a production system to better cope with fluctuating supply and demand. We argue that good flexible process structures

are essentially highly connected graphs, and use the concept of graph expansion (a measure of graph connectivity) to achieve various insights into this design problem.

We analyze the worst-case performance of the flexible design problem under a more general setting, which encompasses a large class of objective functions. We show that whenever demand and supply are balanced and symmetrical, the graph expander structure (a highly connected but sparse graph) is within ϵ-optimality of the fully flexible system, *for all demand scenarios*, although it uses a far smaller number of links. Furthermore, the same graph expander structure works uniformly well for all objective functions in this class. We also generalize this result to the non-symmetrical system, which is more relevant in practice, by introducing the notion of Ψ-expander.

Based on this insight, we develop a simple and easy-to-implement heuristic to design flexible process structures. Numerical results show that this heuristic performs well for a variety of numerical examples previously studied in the literature. Our numerical studies also confirm that process structures with good expansion properties have superior average and worst-case performances.

Acknowledgments

The authors thank the area editor, the associate editor and two anonymous referees for their valuable comments and suggestions that helped improve this paper. This research was supported in part by NUS Academic Research Fund R-314-000-082-112 and National Natural Science Foundation of China No. 70901050.

References

Aksin, O. Z., F. Karaesmen. 2007. Characterizing the performance of process flexibility structures. *Operations Research Letters* **35**(4) 477–484.

Asratian, A., T. Denley, R. Haggkvist. 1998. *Bipartite graphs and their applications*. Cambridge University Press.

Bassalygo, L. A., M. S. Pinsker. 1973. Complexity of an optimum non-blocking switching network with reconnections. *Problemy Informatsii (English Translation in Problems of Information Transmission)* **9** 84–87.

Bassamboo, A., R.S. Randhawa, J.A. Van Mieghem. 2009. Optimal flexibility configurations in newsvendor networks: Going beyond chaining and pairing. *Working paper*.

Benjafaar, S. 2002. Modeling and analysis of congestion in the design of facility layouts. *Management Science* **48**(5) 679–704.

Bish, E., A. Muriel, S. Biller. 2005. Managing flexible capacity in a make-to-order environment. *Management Science* **51** 167–180.

Chou, M. C., G. Chua, C. P. Teo, H. Zheng. 2010. Design for process flexibility: efficiency of the long chain and sparse structure. *Operations Research* **58**(1) 43–58.

de Farias, D. P., B. Van Roy. 2004. On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res.* **29**(3) 462–478.

Graves, S. C., B. T. Tomlin. 2003. Process flexibility in supply chain. *Management Science* **49**(7) 907–919.

Gurumurthi, S., S. Benjaafar. 2004. Modeling and analysis of flexible queueing systems. *Naval Research Logistics* **51** 755–782.

Hopp, W. J., E. Tekin, M. P. Van Oyen. 2004. Benefits of skill chaining in production lines with cross-trained workers. *Management Science* **50**(1) 83–98.

Iravani, S. M., M. P. Van Oyen, K. T. Sims. 2005. Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Science* **51**(2) 151–166.

Jack, E. P, A. S Raturi. 2003. Measuring and comparing volume flexibility in the capital goods industry. *Production and Operations Management* **12**(4) 480.

Jordan, W. C., S. C. Graves. 1995. Principles on the benefits of manufacturing process fexibility. *Management Science* **41**(4) 577–594.

Reingold, O., S. Vadhan, A. Wigderson. 2002. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. *Ann. of Math.* **155** 157–187.

Sarnak, P. 2004. What is an expander? *Notices of the AMS* **51**(7) 762–763.

Sethi, A. K., S. P. Sethi. 1990. Flexibility in manufacturing: a survey. *The International Journal of Flexible Manufacturing Systems* **2** 289–328.

Shi, D., R. L. Daniels. 2003. A survey of manufacturing flexibility: Implications for e-business. *IBM Systems Journal* **42**(3) 414–427.

Van Mieghem, J. A., N. Rudi. 2002. Newsvendor networks: Inventory management and capacity investment with discretionary activities. *Manufacturing & Service Operations Management* **4**(4) 313–335.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Proof of Theorem 4

We start the proof with a roadmap outlining the key steps:

1. We use KKT conditions to characterize $x^*_{i,j}$, for all edge (i,j), the optimal flows between the plant and the product nodes and $U'_j\left(\sum_{l:l\in\mathcal{A}} x^*_{l,j}\right)$, for all plant $j\in\mathcal{B}$, the marginal utility for each plant node j. Using these characteristics, we can partition the plant nodes into groups while those plant nodes in the same group have the same marginal utility. We can then rank the groups in an increasing order of its marginal utility.

2. We focus on the groups with marginal utility less than K, and refer to the set of edges connecting to the union of these groups as $\mathcal{S}_0$.

3. We focus on the set of product nodes adjacent to at least one of the edges in $\mathcal{S}_0$. That is, we focus on $\mathcal{A}\cap\mathcal{S}_0$, and refer to this product set as $\mathcal{T}$. We note that $\Gamma(\mathcal{T})$, the neighbor of $\mathcal{T}$, is in fact the same as $\mathcal{B}\cap\mathcal{S}_0$.

4. We consider three cases: case (a) $\mathcal{T}=\emptyset$, case (b) $\mathcal{T}$ is a *small* subset, and case (c) $\mathcal{T}$ is a *non-small* subset.

5. In case (a), $\mathcal{T}=\emptyset$ implies that the marginal utility for any plant $j\in\mathcal{B}$ is no less than K. Since the utility function of any plant is a concave function with marginal utility less than K when the plant operates beyond its pre-confiured capacity level, we can show that all plants operate under their pre-configured capacity level.

6. In case (b), using the definition of Ψ-expander (Definition 5), the definition of a small subset (Definition 6), and the propoerties of the utility functions, we can show that this case is not possible.

7. In case (c), using the definition of Ψ-expander, the definition of a non-small subset (Definition 6), and the propoerties of the utility functions, we can show that at least Ψ proportion of the pre-configured capacity is utilized in this case.

The details of the proof are in the following.

Consider any given $\tilde{\boldsymbol{D}}=\{\tilde{D}_i\}$. The KKT conditions are the same as the conditions for the symmetrical problem considered in Theorem 1, except that (3) needs to be adjusted slightly as the

following:

$$U_j'\left(\sum_{l\in\mathcal{A}} x_{lj}^*\right) - u_i^* + v_{ij}^* = 0 \ \forall\ (i,j)\in\mathcal{F} \tag{EC.1}$$

Let $\mathcal{S}(\tilde{\boldsymbol{D}}) \triangleq \big\{(i,j) : x_{i,j}^* > 0\big\}$ and $\bar{\mathcal{S}}(\tilde{\boldsymbol{D}}) \triangleq \big\{(i,j) : x_{i,j}^* = 0\big\}$. $\mathcal{S}(\tilde{\boldsymbol{D}})$ can be easily written as a union of connected components $\mathcal{S}_k$, $k = 1,\ldots,h$. The KKT conditions ensure that, for any $k = 1,\ldots,h$,

$$U_j'\left(\sum_{i:i\in\mathcal{A}} x_{i,j}^*\right) = \beta_k,\ \forall j \in \mathcal{B}\cap\mathcal{S}_k,$$

where β_k is a constant. WLOG we can assume that $\beta_1 < \beta_2 < \ldots < \beta_h$, since we can otherwise combine components with identical β_k together.

Let $\mathcal{S}_0 \triangleq \{\cup\mathcal{S}_i : \beta_i < K\}$, $\mathcal{T} \triangleq \mathcal{A}\cap\mathcal{S}_0$, and $\bar{\mathcal{S}}_0 \triangleq \mathcal{S}(\tilde{\boldsymbol{D}})/\mathcal{S}_0$, $\bar{\mathcal{T}} \triangleq \mathcal{A}\cap\bar{\mathcal{S}}_0$.

In the structure $\mathcal{F}$, we note that

$$\Gamma(\mathcal{T}) = \Gamma(\mathcal{A}\cap\mathcal{S}_0) \subseteq \mathcal{B}\cap\mathcal{S}_0. \tag{EC.2}$$

This is because if (EC.2) does not hold, then there exists an edge $(i,j)\in\mathcal{F}$ with $i\in\mathcal{A}\cap\mathcal{S}_k$, for some $\mathcal{S}_k \subseteq \mathcal{S}_0$, but $j \notin \mathcal{B}\cap\mathcal{S}_0$, which implies that either

- $j \in \mathcal{B}\cap\mathcal{S}_m$, for some $\mathcal{S}_m \subseteq \bar{\mathcal{S}}_0 = \mathcal{S}(\tilde{\boldsymbol{D}})/\mathcal{S}_0$, or
- j has a flow of zero; that is, $x_{i,j}^* = 0$ for all $i\in\mathcal{A}$.

But in the first case, the KKT condition (3) ensures that

$$U_j'\left(\sum_{l\in\mathcal{A}} x_{lj}^*\right) - u_i^* \le 0,$$

i.e., $\beta_m \le u_i^* = \beta_k < K$, which is a contradiction. In the second case, since for all $j\in\mathcal{B}$, $U_j(\cdot)$ is a concave function and $U_j(x) = Kx$ when $0 \le x \le C_j$, we can always reallocate one unit of the demand for i to plant j without decreasing the value of $Z_{\mathcal{F}}(\tilde{\boldsymbol{D}})$. Therefore, WLOG, we can exclude the possibility of the second case. From the above arguments, we know that (EC.2) must hold.

On the other hand, it is easy to see that

$$\mathcal{B}\cap\mathcal{S}_0 \subseteq \Gamma(\mathcal{T}). \tag{EC.3}$$

Hence, we have

$$\Gamma(\mathcal{T}) = \mathcal{B} \cap \mathcal{S}_0. \tag{EC.4}$$

Also note that

$$x^*_{ij} = 0, \quad \forall i \in \bar{\mathcal{T}} \text{ and } j \in \Gamma(\mathcal{T}). \tag{EC.5}$$

(EC.5) holds because otherwise, there must exists an arc $(i,j) \in \mathcal{F}$ with $i \in \bar{\mathcal{T}}$, $j \in \Gamma(\mathcal{T})$, and $x^*_{ij} > 0$. In that case, the KKT conditions ensure that $U_j'(\sum_{l\in\mathcal{A}} x^*_{lj}) = u_i \geq K$, which contradicts that $U_j'(\sum_{l\in\mathcal{A}} x^*_{lj}) < K$ for all $j \in \Gamma(\mathcal{T})$.

From (EC.4) and (EC.5), we must have

$$\sum_{i\in\mathcal{A}\cap\bar{\mathcal{S}}_0} \left(\sum_{j\in\mathcal{B}} x^*_{ij}\right) = \sum_{j\in\mathcal{B}\cap\bar{\mathcal{S}}_0} \left(\sum_{i\in\mathcal{A}} x^*_{ij}\right). \tag{EC.6}$$

Similarly, we can see that

$$\sum_{i\in\mathcal{A}\cap\mathcal{S}_0} \left(\sum_{j\in\mathcal{B}} x^*_{ij}\right) = \sum_{j\in\mathcal{B}\cap\mathcal{S}_0} \left(\sum_{i\in\mathcal{A}} x^*_{ij}\right). \tag{EC.7}$$

We now consider three cases:

Case (a): If $\mathcal{T} = \emptyset$, then $\mathcal{S}_0 = \emptyset$ and $\bar{\mathcal{S}}_0 = \mathcal{S}(\tilde{\boldsymbol{D}})$. Note that for all $j \in \mathcal{B}$, either

- $j \in \mathcal{S}(\tilde{\boldsymbol{D}}) \cap \mathcal{B}$, or
- $j \in \bar{\mathcal{S}}(\tilde{\boldsymbol{D}}) \cap \mathcal{B}$.

In the first case, since $\bar{\mathcal{S}}_0 = \mathcal{S}(\tilde{\boldsymbol{D}})$, we have $j \in \bar{\mathcal{S}}_0 \cap \mathcal{B}$. Therefore, $U_j'(\sum_{l\in\mathcal{A}} x^*_{lj}) \geq K$ since $\beta_k \geq K$ for all $\mathcal{S}_k \subseteq \bar{\mathcal{S}}_0$. Also note that $U_j(x)$ is a concave function with $U_j'(x) < K$ when $x > C_j$, thus we have $\sum_{l\in\mathcal{A}} x^*_{lj} \leq C_j$. In the second case, from the definition of $\bar{\mathcal{S}}(\tilde{\boldsymbol{D}})$, it is obvious that $\sum_{l\in\mathcal{A}} x^*_{lj} = 0 \leq C_j$. Hence, combining the above two cases, we conclude that $\sum_{l\in\mathcal{A}} x^*_{lj} \leq C_j$ for all $j \in \mathcal{B}$. That is, all plants operate under their pre-configured capacity level.

Case (b): If $\mathcal{T}$ is a *small* subset, then from (EC.4), (EC.7), the definition of Ψ-expander (Definition 5), and the definition of a small subset (Definition 6), we must have

$$\sum_{j\in\Gamma(\mathcal{T})} \left(\sum_{i\in\mathcal{A}} x^*_{ij}\right) = \sum_{i\in\mathcal{T}} \tilde{D}_i \leq \sum_{i\in\mathcal{T}} \lambda_i^U \mu_i \leq \sum_{j\in\Gamma(\mathcal{T})} C_j.$$

However, since

- $U_j(x)$ is a concave function with $U_j'(x) = K$ when $0 \le x \le C_j$, and
- $U_j'(\sum_{i\in\mathcal{A}} x_{ij}^*) < K$ for all $j \in \Gamma(\mathcal{T})$,

we must have $\sum_{i\in\mathcal{A}} x_{ij}^* > C_j$ for all $j \in \Gamma(\mathcal{T})$, and hence, $\sum_{j\in\Gamma(\mathcal{T})}\left(\sum_{i\in\mathcal{A}} x_{ij}^*\right) > \sum_{j\in\Gamma(\mathcal{T})} C_j$, which is a contradiction. Thus $\mathcal{T}$ cannot be a *small* subset.

Case (c): If $\mathcal{T}$ is a *non-small* subset, then from the definition of a non-small subset (Definition 6), we have $\sum_{i\in\mathcal{T}} \lambda_i^U \mu_i > \Psi \sum_{j\in\mathcal{B}} C_j - \sum_{i\notin\mathcal{T}} \lambda_i^L \mu_i$.

- If $\sum_{i\in\mathcal{T}} \tilde{D}_i \le \sum_{j\in\Gamma(\mathcal{T})} C_j$, then using some of the arguments in Case(b), we can show that $\sum_{j\in\Gamma(\mathcal{T})}\left(\sum_{i\in\mathcal{A}} x_{ij}^*\right) \le \sum_{j\in\Gamma(\mathcal{T})} C_j$ and $\sum_{j\in\Gamma(\mathcal{T})}\left(\sum_{i\in\mathcal{A}} x_{ij}^*\right) > \sum_{j\in\Gamma(\mathcal{T})} C_j$, which is a contradiction.
- If $\sum_{i\in\mathcal{T}} \tilde{D}_i > \sum_{j\in\Gamma(\mathcal{T})} C_j$, then for all $j \in \Gamma(\mathcal{T})$, $\sum_{i\in\mathcal{T}} x_{ij}^* \ge C_j$ since $U_j(\cdot)$ is a non-decreasing concave function with $U_j(x) = Kx$ for $0 \le x \le C_j$ and $U_j'(x) < K$ for $x > C_j$. Since $\sum_{i\in\mathcal{A}} x_{ij}^* \ge \sum_{i\in\mathcal{T}} x_{ij}^*$, we have $\sum_{i\in\mathcal{A}} x_{ij}^* \ge C_j,\ \forall j \in \Gamma(\mathcal{T})$, thus $\sum_{j\in\Gamma(\mathcal{T})} C_j$ is fully utilized.

Note that

$$\sum_{i\in\mathcal{A}} x_{ij}^* \le C_j,\ \forall j \in \mathcal{B} \cap \bar{\mathcal{S}}_0,$$

because $U_j'\left(\sum_{i\in\mathcal{A}} x_{ij}^*\right) \ge K$ for all $j \in \mathcal{B} \cap \bar{\mathcal{S}}_0$. Also note that, by Equation (EC.6), we have

$$\sum_{i\in\bar{\mathcal{T}}} \lambda_i^L \mu_i \le \sum_{i\in\bar{\mathcal{T}}}\left(\sum_{j\in\mathcal{B}} x_{ij}^*\right) = \sum_{j\in\mathcal{B}\cap\bar{\mathcal{S}}_0}\left(\sum_{i\in\mathcal{A}} x_{ij}^*\right) \le \sum_{j\in\mathcal{B}\cap\bar{\mathcal{S}}_0} C_j.$$

Therefore, all the plants in $\mathcal{B} \cap \bar{\mathcal{S}}_0$ operate within its pre-configured capacity C_j, with at least $\sum_{i\in\bar{\mathcal{T}}} \lambda_i^L \mu_i$ capacity utilized. According to the definition of Ψ-expander, we know that

$$\sum_{j\in\Gamma(\mathcal{T})} C_j + \sum_{i\in\bar{\mathcal{T}}} \lambda_i^L \mu_i \ge \Psi \sum_{j\in\mathcal{B}} C_j,$$

hence we have at least Ψ proportion of the pre-configured capacity being utilized. □