

Modularity measures: Concepts, computation, and applications to manufacturing systems

Yue Shao | Victor M. Zavala 

Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin

Correspondence

Victor M. Zavala, Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Dr., Madison, WI 53706, USA.
Email: victor.zavala@wisc.edu

Funding information

NSF CAREER, Grant/Award Number: CBET-1748516

Abstract

We propose a measure to quantify the modularity of industrial production (manufacturing) systems and optimization formulations to compute it. From a manufacturing perspective, we argue that a system is deemed modular if: (a) the equipment units that comprise it form clusters (modules) of dense connectivity (i.e., difficult module assembly tasks are performed off-site), (b) connectivity between modules is sparse (i.e., easy assembly tasks are performed on-site), (c) the number of modules is small, and (d) the module dimensions facilitate transportation. We show that the measure proposed satisfies these requirements and that it can be computed by solving a convex mixed-integer quadratic program. We provide a discussion on advantages and disadvantages of alternative modularity measures used in different scientific and engineering communities. Our results seek to highlight conceptual and computational challenges that arise from the need to define and quantify modularity in a manufacturing context.

KEYWORDS

graph theory, manufacturing, modularity, optimization, organization

1 | INTRODUCTION

Modularization is an organization strategy that is used in living, socio-economic, and industrial systems to facilitate learning and evolution and to cope with complexity.^{1,2} For instance, biological networks and the human body exhibit high modularity.³⁻⁶ This organization structure facilitates specialization of components (e.g., organs and metabolic cycles) and enables management of large numbers of functions. In a modular organization, fundamental components and associated functions are grouped into clusters (modules). Modules have the distinctive feature that coupling between internal components (intra-module) is significantly stronger than coupling across modules (intermodule). Scientists have long argued that modular organization provides flexibility and facilitates evolution because modules can adapt, mature, or disappear without significantly disrupting the entire system. This arrangement also facilitates the management of complexity because tasks and information are refined progressively. Herbert Simon, one of the pioneers of computer and cognitive science, argued that it is rather natural that human-made organizations

(e.g., government institutions and enterprises) also exhibit high degrees of modularity.¹ This is because the human brain processes information and makes decisions in a modular manner.⁶ Modularity provides an indication of the flexibility and maturity of an organization and of the range of functions that it can perform.² Modularity has also been found to facilitate the control of large networks.⁷

Modularity concepts have also been recently explored in the context of industrial production (manufacturing) systems such as chemical processes, energy systems, and infrastructures. Industrial production systems can be built from small-scale and standardized equipment modules that perform well-defined tasks and that are coupled together using well-defined and sparse interfaces.⁸⁻¹⁰ Standardization and size reduction enables mass off-site fabrication and fast transportation and deployment of equipment, which accelerates experimentation and learning and ultimately leads to technology cost reductions.¹¹⁻¹⁶ A celebrated example of this principle is Henry Ford's assembly line.² Modular systems contrast with large and customized systems, which involve lengthy on-site construction phases and difficult transportation (and are thus rarely relocated); these systems also

provide limited experimentation/testing opportunities.¹⁷ Modular systems also enable sequential investment strategies, which provide flexibility to mitigate market and regulatory risk.¹⁸ Small modular systems can also facilitate the processing of geographically dispersed resources that are deemed too expensive to collect and centralize.

Modularization can accelerate investment in technologies such as small nuclear reactors, distributed generators, power electronics, chemical processes, and battery storage systems.^{19–23} Specifically, large industrial facilities (reaching investments of billions of U.S. dollars) might involve slow deployments and risks that few investors are willing to tolerate. On the other hand, modularization provides flexibility in investment size and enables faster deployments that ultimately result in reductions in time-to-market and facilitates financing. Moreover, expansion of production capacity in modular systems can proceed sequentially, which provides a mechanism to hedge against the market and regulatory risk. We can interpret the ability to accelerate and stage investment (and thus a hedge against risk) as a form of built-in logistical temporal flexibility. Modularity can also provide logistical spatial flexibility in the sense that small modules can be easily transported and relocated. This can enable the recovery of resources that are highly distributed and potentially short-lived. As a result, it has been argued that modularization can enable more sustainable systems and circular economies.²⁴ For instance, modular systems can be used to harness natural gas resources that remain stranded at oil production facilities due to limited gas pipeline infrastructure.²⁵ Modular technologies can also be used to recover bio-gas from organic waste generated at animal farms, landfills, and waste-water treatment facilities. It has also been recently observed that modular systems can be strategically placed to exploit space–time electricity price dynamics and with this mitigate risk.²⁶ In this context, module transportability is important from a relocation perspective. For instance, unlike large central systems, modular systems might not be permanently placed at a single location but might be disassembled, relocated, and reassembled at different locations throughout their lifetime based on changes in resource availability, policy, weather, and infrastructure. For instance, a change in government regulations might render a given facility location undesirable or obsolete. Small modules can also be transported back to shops to perform maintenance and can be quickly replaced. As can be seen, space–time logistical flexibility provided by modularity can allow organizations to diversify, mitigate risk, and have a higher likelihood of surviving strong fluctuations of markets, government regulations, and other externalities. On the downside, the flexibility provided by small modular systems often comes at the expense of higher investments and reduced operational efficiency when comparing to large systems. Specifically, economies of scale benefit large systems due to the favorable scaling of throughput with equipment size. Industrial systems will thus likely evolve into a mixed state in which certain tasks are performed in small dispersed modular systems while others are performed in large centralized facilities. This has the potential of inducing a reorganization of production facilities and of entire supply chains. This is particularly the case in chemical processes and power plants.²⁷ This reasoning also indicates that large centralized facilities are operationally efficient but logistically inefficient from an assembly and transportation stand-point.

The concept of modularity is pervasive in science and engineering but, surprisingly, there are few quantifiable measures of modularity. The availability of proper measures is key to enable more systematic analysis, design, and comparison of modular systems. In pioneering work, Newman proposed a modularity measure that quantifies the edge density of a system (represented as a graph) relative to the expected edge density of a random graph.^{5,28} The argument behind this measure is that modular organizations that arise in natural systems are nonrandom. This measure is intuitive and has seen many interesting applications; for instance, this measure has been shown to provide a flexible and powerful tool for the analysis and design of control architectures^{29–34} and for the decomposition of large-scale optimization problems.^{35,36} A powerful generalization of Newman's measure has been proposed in Reference 37 and here it was shown that systems of high modularity are extremum points of a Hamiltonian function. Information-theoretical interpretations of modularity have also been proposed in the literature.³⁸

Unfortunately, the modularity measure proposed by Newman (and its generalizations) does not have an intuitive interpretation from a manufacturing perspective and fails to capture some desirable features arising in this context (e.g., module dimensions). In order to define alternative modularity measures, it is important to highlight that: modularity is not a classification but a measure (i.e., systems have different degrees of modularity). Moreover, one should note that a system with fixed physical connectivity (topology) can have different modular organizations with associated degrees of modularity and that topology dictates the number of alternative modular organizations and their associated modularity value. While these notions are clear from a conceptual point of view, there is significant ambiguity associated with the definition of modularity in manufacturing. In the metal processing industry, for instance, a module is defined as a technically and organizationally limited area of a facility that fulfills a defined task in terms of company-internal or -external salable goods and services.³⁹ In the process industry, a module is defined as an unmodifiable element that provides a dedicated function for the process and is reusable during the planning or realization of modular plants.⁴⁰ In other words, a module is a standardized and self-functioning unit. While these definitions are intuitive, they do not provide means to quantify modularity. Specifically, under these definitions, any equipment unit or an entire facility itself can be a module. Moreover, these definitions fail to capture aspects such as transportability and dimensions.

In this work, we propose measures to quantify the modularity of manufacturing systems and optimization formulations to compute them. We claim that, from a manufacturing perspective, a system is deemed modular if: (a) the equipment units that compose it form clusters (modules) of dense connectivity (i.e., difficult module construction is performed off-site), (b) connectivity between modules is sparse (i.e., easy module assembly is performed on-site), (c) the number of modules is small, and (d) the module dimensions facilitate transportation. In the proposed framework, a facility has a topology that is modeled as a graph. Here, the physical equipment units represent nodes that are coupled together via edges. This representation allows us to borrow concepts and techniques from graph theory. Specifically, from a graph-theoretical perspective, the partitioning of a graph induces a modular organization, as each partition (module) is composed of a set of nodes. For a given

organization, the amount of internal module coupling relative to the coupling between modules is referred to as the *modularity*.⁵ In our approach, the proposed measure is computed for a graph by finding the partition that induces the maximum modularity (given a fixed number of modules). We show that this measure can be computed by solving a convex mixed-integer quadratic program (MIQP). Moreover, we show that the mixed-integer representation allows us to impose additional features such as module dimensions and to identify multiple solutions that give the same level of modularity. We compare the proposed measure against that of Newman (widely used in other scientific communities) to highlight the advantages and disadvantages from a manufacturing perspective. This analysis reveals that the mixed-integer programming formulations proposed can also be used to compute the measure of Newman while handling constraints and that they can be used to find multiple solutions. Moreover, the proposed measure can be used within optimal design formulations and in other applications beyond manufacturing (e.g., design of control architectures and decomposition of large sets of equations).

2 | MEASURES OF MODULARITY

In our framework, we assume that connectivity, number of modules, and dimensions are key features that define the modularity of a system. Connectivity and number of modules dictate the nature and complexity of off-site and on-site assembly tasks while dimensions dictate whether modules are transportable (and thus off-site assembly is possible) and dictate economies of scale. Module connectivity is related to the problem of community detection in networks, which has been widely studied in graph theory.⁴¹ Motivated by this, it seems natural to model a manufacturing system as a graph that is composed of nodes (equipment units) and edges (that connect the units) and use graph theoretical techniques to identify modular organizations. Current techniques available include hierarchical clustering algorithms,^{42,43} k-means clustering,⁴⁴ spectral clustering,⁴⁵ and techniques based on modularity maximization.^{5,28} An excellent review on community detection techniques is provided in Reference 46. In this work, we adopt a modularity maximization approach, as this provides an intuitive approach to analyze and design modular manufacturing systems.

Transportation logistics is a key factor that is often overlooked in modularity studies and that is unique to manufacturing (compared to

other scientific disciplines such as neurology). In particular, a commercially-viable equipment module must be transportable using available infrastructure (e.g., railway and trucks).⁴⁷ As such, module dimensions (length, width, and height) and weight must follow government regulations. For instance, according to the United States regulations for commercial motor vehicles, the maximum width allowed in an interstate highway is 102–130 in. (2.6–3.3 m), the maximum height allowed is 14–16 ft (4.3–4.9 m), the maximum length allowed is 75 ft (22.86 m), and the maximum weight is 44,000 lbs.⁴⁷ Consequently, any system that does not satisfy these limits must be partitioned in order to enable transportation. For example, a distillation system with a diameter of more than 120 in. and height of 100 ft (around 40 trays with 24 in. spacing) must be partitioned to enable transportation and must be assembled on-site. As expected, the larger the dimensions of the system the more partitions that will be needed and the more complex the on-site assembly (Figure 1).

We observe that the different features desired for modular systems might be conflicting. For instance, a small system might be the ideal modular system in that it can be completely assembled off-site (i.e., minimizing on-site assembly tasks), packed in a single module, and transported to its final destination. However, this small system might be inefficient from the perspective of economies of scale. Consequently, one might be willing to modularize only certain components of the system (thus increasing the number of modules but increasing efficiency). As another example, note that one might intentionally prefer a system with a larger number of modules in order to facilitate shop assembly of different types of modules and at different locations.

3 | GRAPH THEORETICAL CONCEPTS

We model a system as an undirected graph $G = (V, E)$ where V is its set of nodes (vertices) and E is its set of edges. We define the number of nodes as $n = |V|$ and the number of edges as $m = |E|$. Connectivity of nodes in G is encoded in the adjacency matrix $A \in \mathbb{R}^{n \times n}$ with entries $A_{i,j}$, $i, j \in V$. We have that $A_{i,j} = 1$ if node i and node j are connected by an edge or $A_{i,j} = 0$ otherwise. We also have that A is symmetric (i.e., $A_{i,j} = A_{j,i}$) and we assume that no self-connections are present (i.e., $A_{i,i} = 0$).

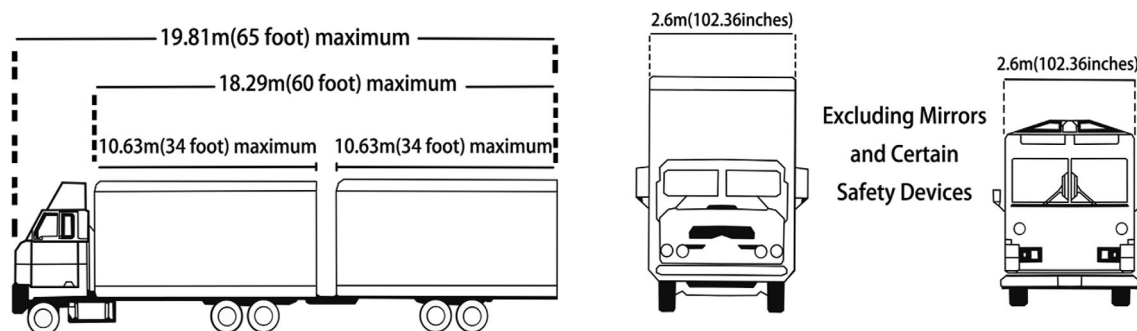


FIGURE 1 United States federal regulations on transportation dimensions⁵⁰

A graph $G = (V, E)$ admits multiple possible modular organizations. A given organization partitions the node set V into a set of modules C and we define the number of modules as $t = |C|$. A module $c \in C$ is a node collection $V_c \subseteq V$ and we have that $\bigcup_{c \in C} V_c = V$ and $V_c \cap V_{c'} = \emptyset, c, c' \in C$ (modules have nonoverlapping elements). We use notation $c(i) \in C, i \in V$ to denote the module that node $i \in V$ belongs to. We define the binary module membership matrix $\delta \in \{0, 1\}^{n \times n}$ with entries:

$$\delta_{ij} = \begin{cases} 1 & \text{if } c(i) = c(j) \\ 0 & \text{otherwise} \end{cases} \quad i, j \in V. \quad (1)$$

In other words, $\delta_{i,j} = 1$ if nodes i, j are in the same module or $\delta_{i,j} = 0$ otherwise. If all nodes are in the same module we have that $\sum_{i \in V} \sum_{j \in V} \delta_{ij} = n \cdot n$; on the other hand, if all nodes are in separate modules, we have that $\sum_{i \in V} \sum_{j \in V} \delta_{ij} = n$ because $\delta_{i,j} = 1$ for $i = j$ and $\delta_{i,j} = 0$ for $i \neq j$ (i.e., the membership matrix is the identity matrix). An important property of the membership matrix is that $\text{rank}(\delta) = t$. In other words, the number of modules equals the number of linearly independent columns (or rows) of δ . As can be seen, the membership matrix δ encodes all relevant information associated with a given modular organization.

To illustrate the relationship between the rank of the membership matrix and the number of modules, suppose that we have a graph with four nodes $V = \{a, b, c, d\}$ and $C = \{1, 2\}$ modules (and thus $t = 2$). Assume that nodes a and b are in the same module and nodes c and d are in the another module. Therefore, we have $\delta_{a,b} = \delta_{c,d} = 1$, and $\delta_{a,d} = \delta_{c,b} = 0$. We define the columns of δ corresponding to nodes a and c as δ_a and δ_c , and we would like to show that these columns are linearly independent. Equivalently, we want to prove that the only solution to the linear system

$$S_1 \delta_{a,b} + S_2 \delta_{c,b} = 0 \quad (2a)$$

$$S_1 \delta_{a,d} + S_2 \delta_{c,d} = 0 \quad (2b)$$

is $S_1 = S_2 = 0$. Upon substitution of $\delta_{i,j}$ in the above equations we obtain that $S_1 = S_2 = 0$ and we thus have $\text{rank}(\delta) = t = 2$.

The partitioning of the graph $G = (V, E)$ into modules induces an organization with a given connectivity inside modules (intramodules) and between modules (intermodules). We measure modularity of an organization as the density of internal module edges relative to the total number of edges. This measure is known as the *graph coverage* and can be computed as:

$$\begin{aligned} \text{cov}(\delta) &= \frac{1}{m} \sum_{c \in C} |E(c)| \\ &= \frac{1}{2m} \sum_{i,j \in V} \delta_{ij} A_{ij} \end{aligned} \quad (3)$$

where $|E(c)|$ denotes the number of edges in module $c \in C$. The factor $1/2$ eliminates the repeated counting of edges in the adjacency matrix A .

We define the modularity measure of system G as:

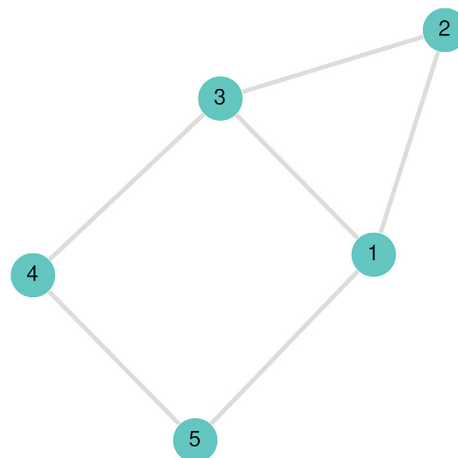


FIGURE 2 Example graph $G = (V, E)$ used to illustrate graph theoretical concepts [Color figure can be viewed at wileyonlinelibrary.com]

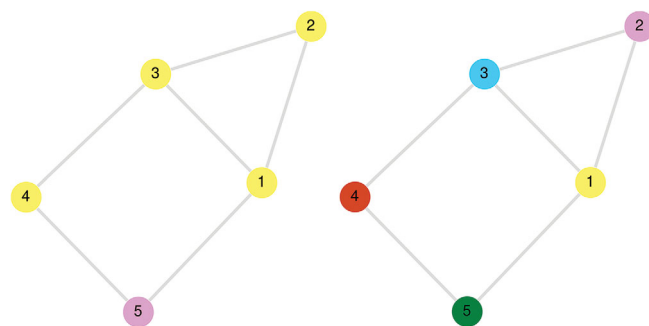


FIGURE 3 Modular organizations for example graph with $t = 2$ (left) and $t = 5$ (right) [Color figure can be viewed at wileyonlinelibrary.com]

$$M_t := \max_{\delta} \text{cov}(\delta) \quad (4a)$$

$$\text{s.t. } \text{rank}(\delta) \geq t. \quad (4b)$$

In other words, the modularity measure is computed by solving a rank-constrained optimization problem. We note that $\sum_{i,j \in V} \delta_{ij} A_{ij}$ is at least zero and at most $2m$ and thus $\text{cov}(\delta) \in [0, 1]$. The $\text{cov}(\delta) = 1$ case corresponds to $\sum_{i,j \in V} \delta_{ij} A_{ij} = 2m$ and occurs when $\delta_{i,j} = 1$ for all $i, j \in V$ (all nodes are in one module and thus $t = 1$). The $\text{cov}(\delta) = 0$ case corresponds to $\sum_{i,j \in V} \delta_{ij} A_{ij} = 0$ and occurs when $\delta_{i,j} = 0$ for all $i, j \in V$ such that $A_{i,j} = 1$. For the maximum possible rank $n = t = \text{rank}(\delta)$, this occurs when $\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ (the membership matrix is the identity matrix). In general, we have that $\text{cov}(\delta)$ is large when connectivity between modules is sparse (connectivity inside modules is dense) and we have that $\text{cov}(\delta)$ is small when the connectivity between modules is dense (inside modules is sparse).

Consistent with the definition of graph coverage, the maximum possible value for M_t is achieved when all nodes are contained in one module ($t = 1$). Consequently, for connected graphs, the value of M_1 is always one. In this case, the modularity measure is given by the unconstrained problem $M_1 = \max_{\delta} \text{cov}(\delta)$. Restricting the number of

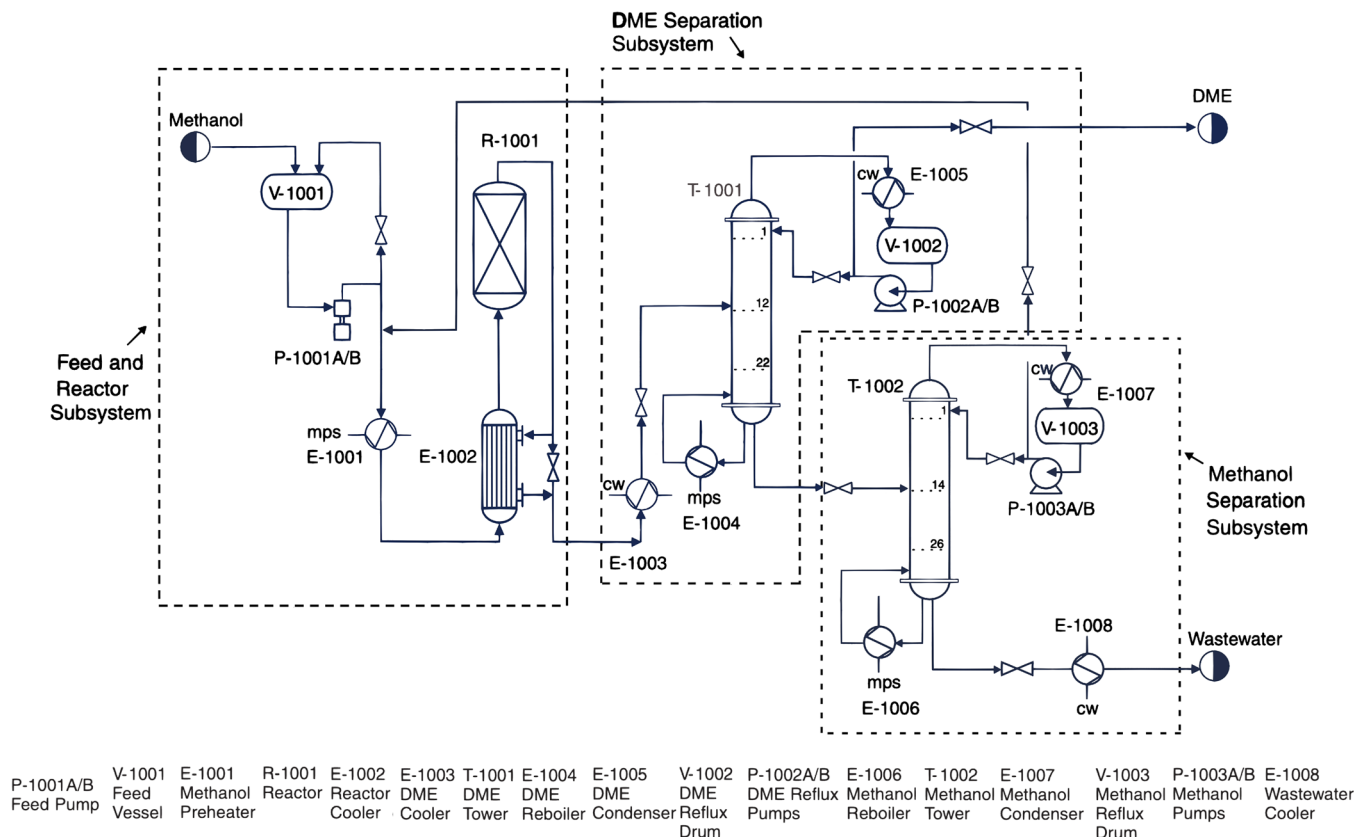


FIGURE 4 Flow diagram and subsystems for dimethyl-ethyl (DME) process (adapted from Reference 48) [Color figure can be viewed at wileyonlinelibrary.com]

modules to any value $t > 1$ forces placement of edges outside modules and thus decreases the modularity measure M_t . In the limit when we restrict $t = n$, we obtain the minimum modularity M_n (corresponding to the case in which every node is in a different module). The proposed modularity measure M_t thus naturally captures trade-offs between graph connectivity and number of modules.

The proposed definition of modularity is intuitive from a manufacturing perspective but alternative definitions exist in the literature (particularly in scientific applications). In Supporting Information we provide a perspective on alternative definitions along with their advantages and limitations.

4 | ILLUSTRATIVE EXAMPLE

We use a simple graph (see Figure 2) with $n = 5$ nodes and $m = 6$ edges to illustrate the concepts. The adjacency matrix A of this graph is:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

The membership matrices δ for two modular organizations with $t = 2$ and $t = 5$ (see Figure 3) are:

$$\delta = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \delta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix on the left has $\text{rank}(\delta) = t = 2$ and is a solution of Problem (4) with $t = 2$ and modularity measure is $M_2 = 4/6$ (coverage $\text{cov}(\delta) = 4/6$). Because this is a solution to Problem (4), any alternative configuration with $t = 2$ must have $M_2 \leq 4/6$. Upon inspection, one can indeed see that the number of internal module edges is four and the total number of edges is $m = 6$. The identity matrix on the right indicates that each node belongs to a module and thus $\text{rank}(\delta) = t = 5$ and is the solution of Problem (4) with $t = 5$ and $M_5 = 0$ (coverage is $\text{cov}(\delta) = 0$).

5 | OPTIMIZATION FORMULATIONS FOR THE MODULARITY MEASURE

We proceed to show that the proposed modularity measure can be formulated as a MIQP. To motivate our discussion, we define the binary variable matrix $x \in \{0, 1\}^{n \times n}$ with entries $x_{ij} = (1 - \delta_{ij})$. The

unconstrained modularity measure can be computed by using the mixed-integer linear program (MILP):

$$\max_x \frac{1}{2m} \sum_{i,j} A_{ij} (1 - x_{ij}) \quad (5a)$$

$$\text{s.t. } x_{ij} \leq x_{i,k} + x_{k,j}, \quad j > i, i, j, k \in V \quad (5b)$$

$$x_{i,i} = 0, i \in V \quad (5c)$$

$$x_{i,j} = x_{j,i}, i, j \in V. \quad (5d)$$

The first constraint enforces the logic that, if i and j are in the same module and i and k are in the same module, then j and k must be in the same module. The second and third constraints capture basic logic that follows from the definition of the membership matrix. This formulation highlights intuitive connections between modularity and mixed-integer formulations. Unfortunately, this MILP formulation does not offer direct control on the number of modules (which is needed to compute measure M_t).

To obtain direct control on the number of modules, we propose a MIQP programming formulation. Here, we define a module (partition) set $C = \{1, \dots, t\}$ with dimension $t \leq n$. We define a binary variable matrix $x \in \{0, 1\}^{n \times t}$ with entries $x_{i,k} = 1$ if node $i \in V$ is in module $k \in C$ and $x_{i,k} = 0$ otherwise. Importantly, under these definitions, we have that:

$$\delta_{ij} = \sum_{k \in C} x_{i,k} x_{j,k}, \quad i, j \in V. \quad (6)$$

Because of this, the modularity measure M_t can be computed by using the MIQP:

$$\max_x \frac{1}{2m} \sum_{i,j \in V} A_{ij} \sum_{k \in C} x_{i,k} x_{j,k} \quad (7a)$$

$$\text{s.t. } \sum_{k \in C} x_{i,k} = 1, \quad i \in V \quad (7b)$$

$$\sum_{i \in V} x_{i,k} \geq 1, \quad k \in C \quad (7c)$$

The first constraint enforces the logic that a node can only belong to one module while the second constraint ensures that at least one node is assigned to each module. The MIQP formulation is expected to be more computationally intensive than the MILP formulation but it captures the features needed (i.e., enforces the rank constraint). For simplicity in the discussion, we transform the MIQP into a minimization problem with objective $-\frac{1}{2m} \sum_{i,j \in V} A_{ij} \sum_{k \in C} x_{i,k} x_{j,k}$.

6 | CONVEXIFICATION OF MIQP FORMULATION

We have found that the MIQP (in minimization form) can be cast as a convex MIQP, which is solvable by modern solvers. To see this, we

define the variable vector $x_k = (x_{1,k}, x_{2,k}, \dots, x_{n,k})$, $k \in C$ and note that we can rewrite the objective function as:

$$\begin{aligned} \frac{1}{2m} \sum_{i,j \in V} A_{ij} \sum_{k \in C} x_{i,k} x_{j,k} &= \frac{1}{2m} \sum_{k \in C} \sum_{i,j \in V} x_{i,k} A_{ij} x_{j,k} \\ &= \frac{1}{2m} \sum_{k \in C} x_k^T A x_k \\ &= \frac{1}{2m} \text{vec}(x)^T H \text{vec}(x) \end{aligned} \quad (8)$$

where H is a block-diagonal matrix of the form:

$$H = \begin{bmatrix} A & & \\ & \ddots & \\ & & A \end{bmatrix} \quad (9)$$

and $\text{vec}(x) = (x_1, x_2, \dots, x_t)$. We note that H is indefinite because the adjacency matrix A is indefinite. However, we note that the entries of $\text{vec}(x)$ are all binary at any feasible solution and thus $\text{vec}(x)^T e = \text{vec}(x)^T \text{vec}(x)$ holds (e is a vector of ones of the same dimension as $\text{vec}(x)$). Consequently, we can write the objective function in the equivalent form:

$$\frac{1}{2m} \text{vec}(x)^T H \text{vec}(x) = \frac{1}{2m} \left(\text{vec}(x)^T (H + I \rho) \text{vec}(x) - \rho \text{vec}(x)^T e \right). \quad (10)$$

for any positive $\rho \in \mathbb{R}_+$ and where I is the identity matrix. This equivalence follows from:

$$\begin{aligned} \frac{1}{2m} \text{vec}(x)^T H \text{vec}(x) &= \frac{1}{2m} \left(\text{vec}(x)^T (H + I \rho) \text{vec}(x) - \rho \text{vec}(x)^T I \text{vec}(x) \right) \\ &= \frac{1}{2m} \left(\text{vec}(x)^T (H + I \rho) \text{vec}(x) - \rho \text{vec}(x)^T e \right) \\ &= \frac{1}{2m} \left(\text{vec}(x)^T H \text{vec}(x) + \rho \text{vec}(x)^T e - \rho \text{vec}(x)^T e \right) \\ &= \frac{1}{2m} \text{vec}(x)^T H \text{vec}(x). \end{aligned} \quad (11)$$

As a result, we can always make the coefficient matrix of the MIQP $(H + I \rho)$ positive definite without affecting the solution and thus make the problem solvable using state-of-the-art solvers. The most obvious choice for ρ would be to use the smallest eigenvalue of H (the smallest eigenvalue of A).

7 | MODELING EXTENSIONS AND OTHER APPLICATIONS

Mixed-integer programming formulations offer flexibility to impose requirements that might be of interest from a manufacturing perspective. For instance, we consider the extended formulation:

$$\max_x \frac{1}{2m} \sum_{i,j \in V} A_{ij} \sum_{k \in C} x_{i,k} x_{j,k} \quad (12a)$$

$$\text{s.t. } \sum_{k \in C} x_{i,k} = 1, \quad i \in V \quad (12b)$$

$$\sum_{i \in V} x_{i,k} \geq 1, \quad k \in C \quad (12c)$$

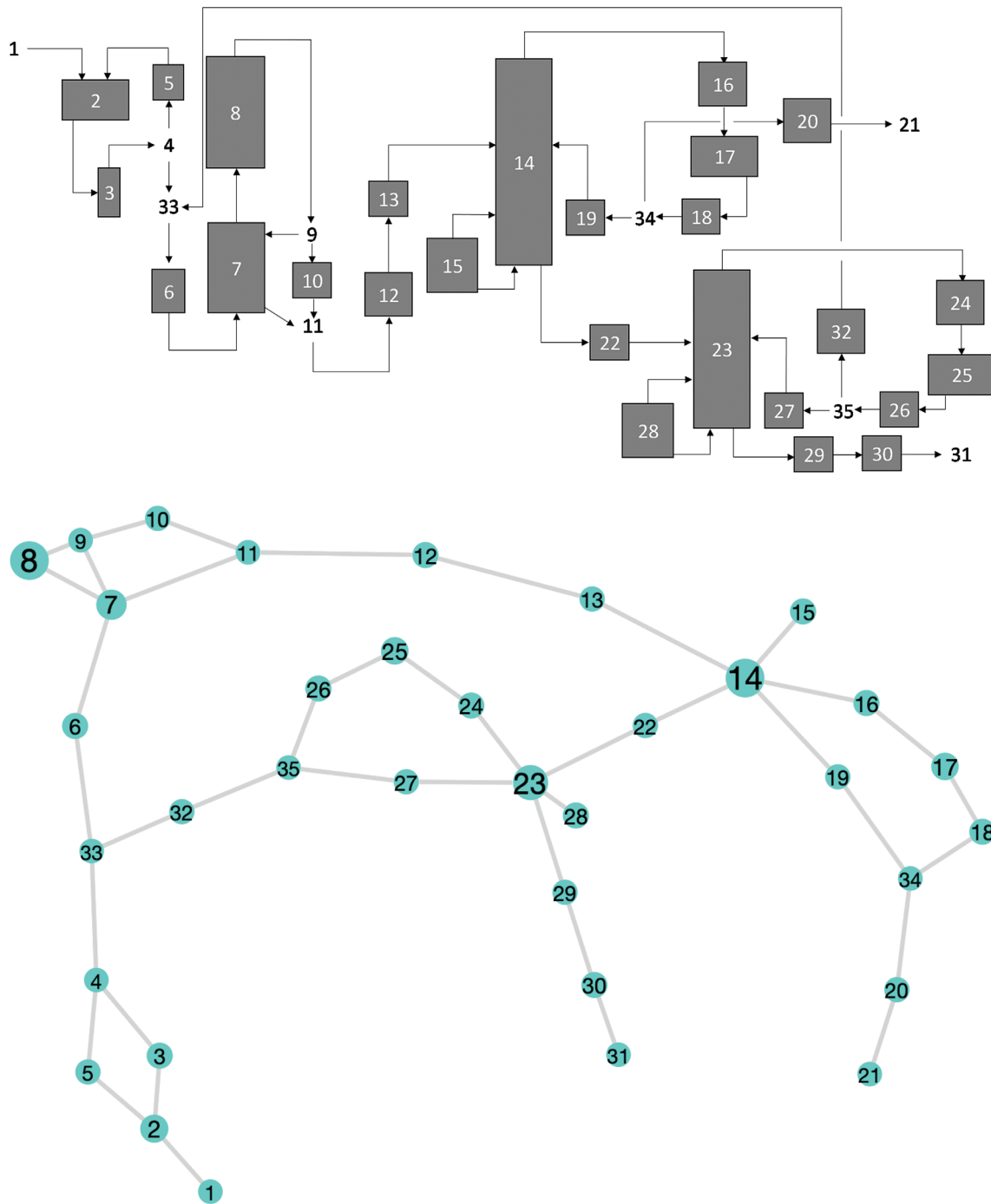


FIGURE 5 Block diagram representation (top) and graph representation (bottom) for DME process [Color figure can be viewed at wileyonlinelibrary.com]

$$\underline{D}_k \leq \sum_{i \in V} x_{i,k} D_i \leq \bar{D}_k, \quad k \in C \quad (12d)$$

Here, the last constraint imposes module feature constraints. The quantity $D_i \in \mathbb{R}_+$ denotes the feature of each node and $\underline{D}_k, \bar{D}_k \in \mathbb{R}_+$ are lower and upper bounds for the features. This constraint can be used to enforce different module features such as weight, height, and

number of nodes in a module. For instance, the number of nodes in a module can be controlled by using the constraint:

$$\underline{D}_k \leq \sum_{i \in V} x_{i,k} \leq \bar{D}_k, \quad k \in C. \quad (13)$$

The proposed formulation can also be extended to impose logic constraints to force/prevent nodes from being in the same modules

Node	Equipment	Dimension	Node	Equipment	Dimension
1	Feed	0	19	Valve	1
2	V-1001	5	20	Valve	1
3	P-1001A/B	2	21	Product	0
4	Flow Junction	0	22	Valve	1
5	Valve	1	23	T-1002	15
6	E-1001	2	24	E-1007	2
7	E-1002	8	25	V-1003	4
8	R-1001	20	26	P-1003A/B	2
9	Flow Junction	0	27	Valve	1
10	Valve	1	28	E-1006	2
11	Flow Junction	0	29	Valve	1
12	E-1003	2	30	E-1008	2
13	Valve	1	31	Product	0
14	T-1001	20	32	Valve	1
15	E-1004	2	33	Flow Junc	0
16	E-1005	2	34	Flow Junc	0
17	V-1002	4	35	Flow Junc	0
18	P-1002A/B	2			

TABLE 1 Node labels and dimensions for each node in the DME process

TABLE 2 Modularity measures and number of alternative configurations for DME process obtained with rank constraints (without dimension constraints)

Rank (t)	Modularity measure (M_t)	Number of configurations	Solution time (s)
1	1	1	14.69
2	0.975	18	14.67
3	0.95	216	14.77
4	0.925	2016	15.68
5	0.9	15,120	15.16
6	0.875	90,720	15.10

and can be extended to identify multiple organizations that lead to the same modularity measure (e.g., by using no-good cuts). In fact, modern mixed-integer solvers can compute all solutions that give the same optimal objective value.

We highlight that the proposed modularity measure and MIQP formulation can be used in other applications that go beyond manufacturing. For instance, these tools can be used to identify optimal configurations for control architectures and optimal decomposition strategies for optimization problems.^{36,46} In this context, constraints on the number and size of modules can be used to create balanced configurations (e.g., to handle computational load balancing issues).

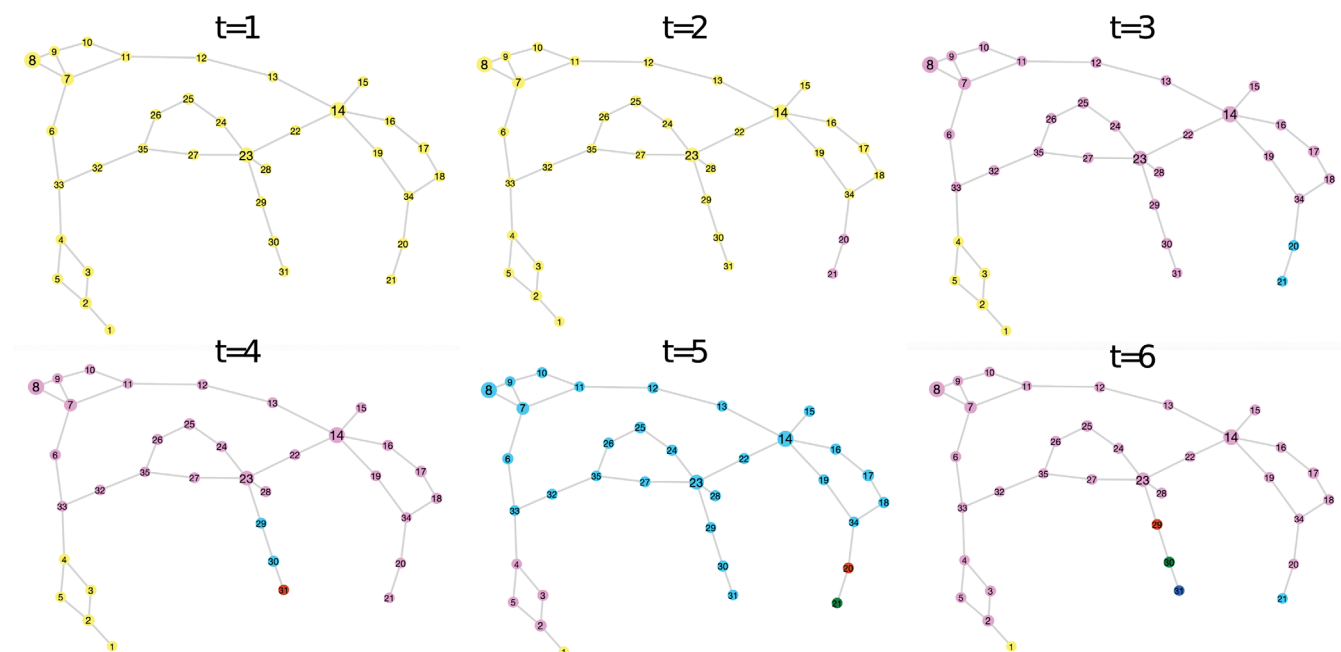


FIGURE 6 Sample modular configurations for DME process for $t \in [1, 6]$ [Color figure can be viewed at wileyonlinelibrary.com]

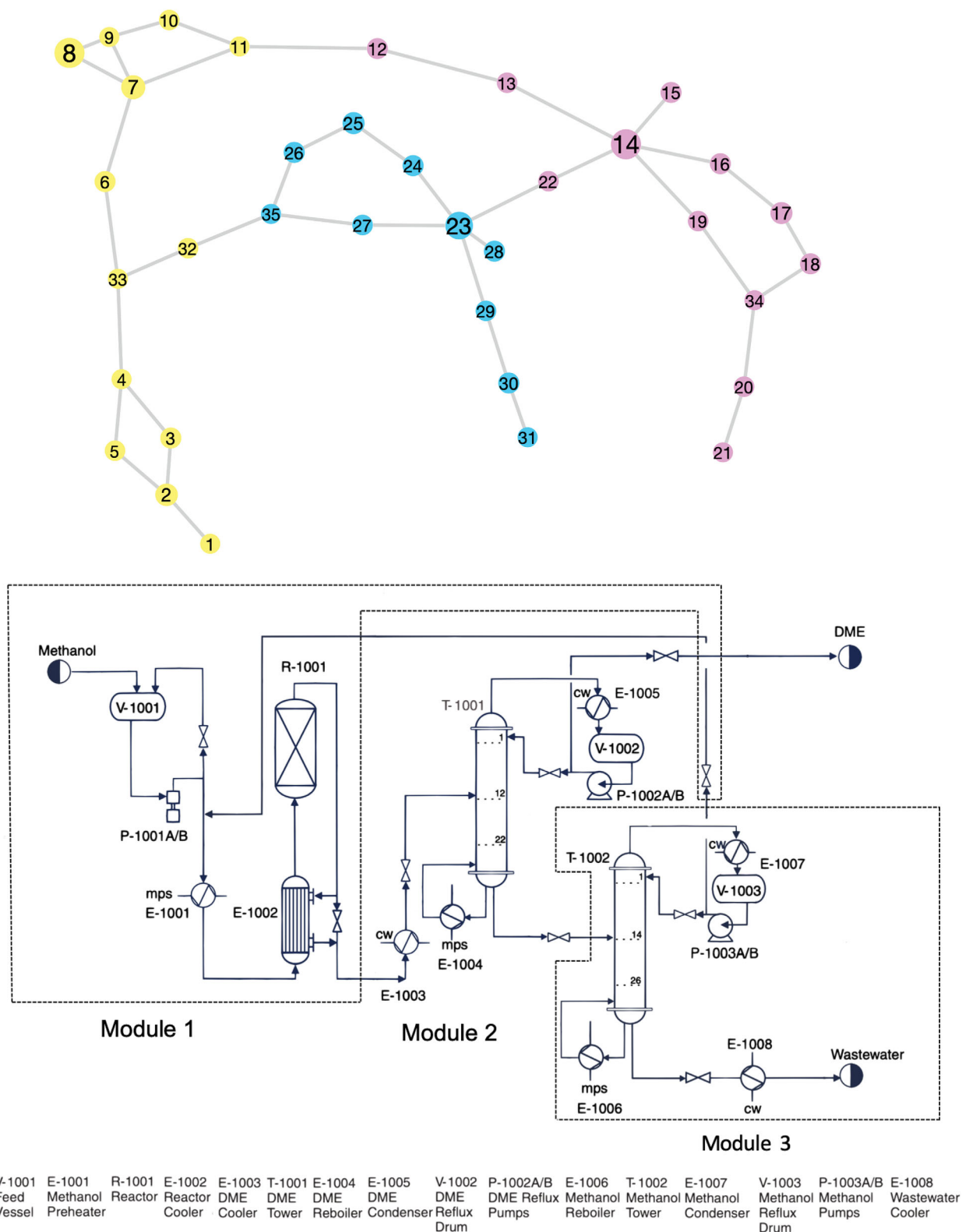


FIGURE 7 Modular configuration for baseline DME process obtained under dimension constraints (adapted from Reference 48). Graph representation (top) and corresponding flowsheet (bottom) [Color figure can be viewed at wileyonlinelibrary.com]

It is important to emphasize that modularity is directly associated to the partitioning of a graph and, as such, there is no analytical representation for such a measure. In other words, one needs to specify the

partition of the graph first and then compute the modularity measure. Because of this, modularity needs to be expressed as an optimization that implicitly finds a partition of maximum modularity.

8 | CASE STUDY

We use the proposed modularity measure and MIQP formulation to identify modular configurations for a dimethyl-ethyl (DME) production process from methanol.⁴⁸ Methanol is an intermediate product during the production of DME from natural gas and thus small modular DME plants can provide a potential pathway to help recover billions of cubic feet of natural gas that are currently stranded and flared. The DME process is intuitively partitioned by practitioners into three functional subsystems: the feed and reactor section, the DME purification section, and the methanol separation and the recycle section. The process flow diagram (PFD) and the subsystems are shown in Figure 4. We created a block and graph representation for the process. To do so, we represent each equipment unit and junction as a node and each flow connection as an edge. The block and graph representations are shown in Figure 5. We use node dimension as a feature that affects the system modularity (i.e., this affects transportation). Labels and dimensions for the nodes are presented in Table 1.

We first computed the unconstrained modularity measure while ignoring rank (number of modules) and dimension constraints. As expected, the solution of this problem gives a modularity measure $M_1 = 1$ (i.e., all nodes are assigned to one module). We then computed the modularity measure by spanning the range $t \in [1, 6]$. The results are summarized in Table 2 and a visualization of each configuration is

presented in Figure 6. We can see that, as the number of modules increases, the modularity measure decreases from $M_1 = 1$ to $M_6 = 0.875$. We thus have that, for a configuration with $t = 6$, 87.5% of the edges are inside the modules while 12.5% connect the modules (the configuration has sparse intermodule coupling).

For every value of t , we computed all possible equivalent configurations (solutions that give the maximum value of M_t). We do this in order to highlight that multiple configurations can give the same modularity measure. We found that the number of alternative solutions increases sharply with the increasing number of modules. This indicates that degeneracy increases with the number of modules and highlights the combinatorial nature of the problem. This also indicates that there is significant flexibility to find configurations that satisfy additional requirements (such as dimension constraints).

We computed the modularity measure by considering dimension constraints (but ignore rank constraints). Here, we explore the impact of scaling-up the process and set the minimum dimension of each module to $\underline{D}_k = 20$ and the maximum dimension $\bar{D}_k = 40$. A visualization of the modular configuration is shown in Figure 7 and the associated node-module membership is shown in Table 3. Here, we also report the module dimensions $\sum_{i \in V} x_{i,k} D_i$ for all $k \in C$. We observe that the dimension constraints induce an organization with $t = 3$ modules and the associated modularity measure is $M_3 = 0.925$ (only 7.5% of the edges connect modules). Interestingly, we can see that the resulting modular organization is the same as the functional organization shown in Figure 4 (with the exception of valves). In fact, if we found that the modularity measure that result from the functional organization of Figure 4 is 0.925 and is thus optimal. This highlights that practitioners use natural logic to modularize systems and that the proposed modularity measure is intuitive.

We explored the effect of scaling up and down on the modularity measure by scaling the equipment unit dimensions. Scaling results are summarized in Table 4. We recall that the baseline measure value is

TABLE 3 Modular configuration for DME process obtained under dimension constraints

Module	Nodes	Module dimension
1	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 32, 33]	40
2	[12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 34]	36
3	[23, 24, 25, 26, 27, 28, 29, 30, 31, 35]	29

TABLE 4 Effect of DME process scaling on modularity measure

Scale	Measure (M_t)	Rank (t)	Solution time (s) with Gurobi	Solution time (s) with convexification
0.3	1.0	1	15.59	16.91
0.5	0.95	2	15.25	16.17
1 (baseline)	0.925	3	15.36	17.67
1.2	0.875	4	15.80	16.48
1.5	0.775	5	15.37	16.73
1.9	0.75	6	16.05	17.01

TABLE 5 Comparison of the number of solutions for problems with and without dimension constraints (DCs)

Rank (t)	Measure (M_t) without DCs	Number of configurations without DCs	Measure (M_t) with DCs	Number of configurations with DCs
3	0.950	216	0.925	78
4	0.925	2016	0.875	384
5	0.900	15,120	0.775	1920

$M_3 = 0.925$. As expected, we observe that the number of modules increases and the modularity measure decreases as we scale up the process. By scaling the equipment units up by 20% the measure decreases to $M_2 = 0.875$. The modularity measure achieves its ideal value of $M_1 = 1$ when the baseline process is scaled down by 30%. This highlights that the modularity measure proposed is consistent and that dimension constraints can also be used to implicitly control the number of modules.

The MIQPs were solved using the Gurobi (version 0.6.0) and were implemented in the Julia-based JuMP modeling framework. We use GraphPlot and LightGraphs for graph manipulation and visualization. All nodes needed to reproduce the results can be found in <https://github.com/zavalab/JuliaBox/tree/master/ModularityMeasures>. We solved the MIQP problems by convexifying them directly. To do so, the minimum eigenvalue of the adjacency matrix A is -2.62 and thus we used $\rho = 3$. We highlight that Gurobi can also automatically convexify the problem (convexification by the user is not needed). We confirmed that both approaches give the same solutions (Gurobi gives slightly better times). The solution times obtained are in the range of 14–17 s (these are reported in Table 4).

We computed the number of solutions for problems with dimension constraints for $t = 3$, $t = 4$, and $t = 5$ and compared against the number of solutions obtained without dimension constraints. The results are summarized in Table 5. We can see that for all cases, the number of solutions is drastically reduced when dimensional constraints are added. This highlights the importance of enforcing additional module features to mitigate the natural degeneracy of modularity measures. In particular, other modularity measures used in the scientific literature, such as that of Newman, are degenerate (i.e., different organizations give the same modularity measure) and this degeneracy can introduce ambiguity in the analysis. Mixed-integer programming approaches allow us to systematically explore the set of feasible solutions.

9 | CONCLUSIONS AND FUTURE WORK

In this work, we propose a modularity measure to guide the analysis, comparison, and design of modular systems. The proposed measure seeks to capture desirable features from a manufacturing perspective. We show that the modularity measure can be computed by solving a MIQP and that this MIQP can be convexified to enable efficient solutions with state-of-the-art solvers. Moreover, this formulation allows us to capture logical constraints associated with module dimensions and node-module membership restrictions. As part of future work, we are interested in using the proposed measure to guide the synthesis of systems with desired modularity properties. Moreover, we are interested in exploring computational strategies to analyze large-scale graphs.

There exist interesting synergies of the modular design principles discussed in this work with modular design principles of control architectures.⁴⁹ This is because in both cases one is implicitly seeking to minimize the degree of interaction between modules. Using the

proposed measure to understand the interplay between modular design and control is an interesting topic of future work.

ACKNOWLEDGMENT

Victor M. Zavala acknowledges funding from the NSF CAREER award CBET-1748516.

ORCID

Victor M. Zavala  <https://orcid.org/0000-0002-5744-7378>

REFERENCES

1. Simon HA. The architecture of complexity. *Proc Am Philos Soc.* 1962; 106(6):467-482.
2. Langlois RN. Modularity in technology and organization. *J Econ Beh Org.* 2002;49(1):19-37.
3. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. *Science.* 2002; 297(5586):1551-1555.
4. Meunier D, Lambiotte R, Bullmore ET. Modular and hierarchically modular organization of brain networks. *Front Neurosci.* 2010;4:200.
5. Newman ME. Modularity and community structure in networks. *Proc Natl Acad Sci.* 2006;103(23):8577-8582.
6. Meunier D, Lambiotte R, Fornito A, Ersche KD, Bullmore ET. Hierarchical modularity in human brain functional networks. *Front Neuroinform.* 2009;3:37.
7. Constantino PH, Tang W, Daoutidis P. Topology effects on sparse control of complex networks with laplacian dynamics. *Sci Rep.* 2019;9 (1):9034.
8. Seifert T, Sievers S, Bramsiepe C, Schembecker G. Small scale, modular and continuous: a new approach in plant design. *Chem Eng Process: Process Intens.* 2012;52:140-150.
9. Bramsiepe C, Sievers S, Seifert T, et al. Low-cost small scale processing technologies for production applications in various environments mass produced factories. *Chem Eng Process: Process Intens.* 2012;51:32-52.
10. Baldea M, Edgar TF, Stanley BL, Kiss AA. Modular manufacturing processes: status, challenges, and opportunities. *AIChE J.* 2017;63(10): 4262-4272.
11. Lier S, Grünewald M. Net present value analysis of modular chemical production plants. *Chem Eng Technol.* 2011;34(5):809-816.
12. Lier S, Wörsdörfer D, Grünewald M. Transformable production concepts: flexible, mobile, decentralized, modular, fast. *ChemBioEng Rev.* 2016;3(1):16-25.
13. Rogers G, Bottaci L. Modular production systems: a new manufacturing paradigm. *J Intell Manuf.* 1997;8(2):147-156.
14. Tatum C. Improving constructibility during conceptual planning. *J Constr Eng Manag.* 1987;113(2):191-207.
15. Roy S. Consider modular plant design. *Chem Eng Proc.* 2017;28-31.
16. Hesler WE. Modular design—where it fits. *Chem Eng Proc.* 1990;76-80.
17. Wells DD. Movement key to pre-fab module use. *Oil Gas J.* 1979;77 (42):148-168.
18. Hagspiel V, Huisman KJ, Kort PM. Volume flexibility and capacity investment under demand uncertainty. *Int J Prod Econ.* 2016;178: 95-108.
19. Berthélemy M, Rangel LE. Nuclear reactors' construction costs: The role of lead-time, standardization and technological progress. *Energy Policy.* 2015;82:118-130.
20. Chakraborty S, Kramer B, Kroposki B. A review of power electronics interfaces for distributed energy systems towards achieving low-cost modular design. *Renew Sustain Energy Rev.* 2009;13(9):2323-2335.
21. DiPippo R. Small geothermal power plants: design, performance and economics. *GHC Bull.* 1999;20(2):1-8.

22. Rothwell G. A real options approach to evaluating new nuclear power plants. *Energy J.* 2006;27(1):37-53.
23. Dong L, Liu H, Riffat S. Development of small-scale and micro-scale biomass-fuelled CHP systems-A literature review. *Appl Therm Eng.* 2009;29(11):2119-2126.
24. Seliger G, Zettl M. Modularization as an enabler for cycle economy. *CIRP Annals-Manuf Technol.* 2008;57(1):133-136.
25. Davis N. Natural gas flaring in North Dakota has declined sharply since 2014. <https://www.eia.gov/todayinenergy/detail.php?id=26632>. 2016. Accessed: July 5, 2017.
26. Shao Y, Zavala VM. Space-time dynamics of electricity markets incentivize technology decentralization. *Comp Chem Eng.* 2019;127:31-40.
27. Peters MS, Timmerhaus KD, West RE, Timmerhaus K, West R. *Plant Design and Economics for Chemical Engineers*. Vol 4. New York: McGraw-Hill; 1968.
28. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E.* 2004;69:026113.
29. Jogwar SS, Daoutidis P. Community-based synthesis of distributed control architectures for integrated process networks. *Chem Eng Sci.* 2017;172:434-443.
30. Moharir M, Pourkargar DB, Almansoori A, Daoutidis P. Graph representation and distributed control of diffusion-convection-reaction system networks. *Chem Eng Sci.* 2019;204:128-139.
31. Daoutidis P, Tang W, Jogwar SS. Decomposing complex plants for distributed control: Perspectives from network theory. *Comp Chem Eng.* 2018;114:43-51.
32. Tang W, Allman A, Pourkargar DB, Daoutidis P. Optimal decomposition for distributed optimization in nonlinear model predictive control through community detection. *Comp Chem Eng.* 2018;111:43-54.
33. Pourkargar DB, Almansoori A, Daoutidis P. Comprehensive study of decomposition effects on distributed output tracking of an integrated process over a wide operating range. *Chem Eng Res Des.* 2018;134:553-563.
34. Pourkargar DB, Moharir M, Almansoori A, Daoutidis P. Distributed estimation and nonlinear model predictive control using community detection. *Ind Eng Chem Res.* 2019;58(30):13495-13507.
35. Moharir M, Kang L, Daoutidis P, Almansoori A. Graph representation and decomposition of ODE/hyperbolic PDE systems. *Comp Chem Eng.* 2017;106:532-543.
36. Allman A, Tang W, Daoutidis P. DeCODE: a community-based algorithm for generating high-quality decompositions of optimization problems. *Optim Eng.* 2019;20(4):1067-1084.
37. Reichardt J, Bornholdt S. Statistical mechanics of community detection. *Phys Rev E.* 2006;74(1):016110.
38. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci.* 2008;105(4):1118-1123.
39. Wiendahl H-P, Nofen D, Carl Hanser FB. *Planung Modularer Fabriken*. Munich, Germany: Carl Hanser Verlag GmbH & Co. KG; 2005.
40. Hohmann L, Kössl K, Kockmann N, Schembecker G, Bramsiepe C. Modules in process industry - A life cycle definition. *Chem Eng Process: Process Intens.* 2017;111:115-126.
41. Fortunato S. Community detection in graphs. *Phys Rep.* 2010;486(3):75-174.
42. Trevor Hastie RT, Friedman J. *The Elements of Statistical Learning*. New York, NY: Springer-Verlag; 2009.
43. Heo S, Marvin WA, Daoutidis P. Automated synthesis of control configurations for process networks based on structural coupling. *Chem Eng Sci.* 2015;136:76-87.
44. Rattigan MJ, Maier M, Jensen D. Graph clustering with network structure indices. Proceedings of the 24th International Conference on Machine Learning, ICML '07; 2007; New York, NY: ACM; pp. 783-790.
45. Donath WE, Hoffman AJ. Lower bounds for the partitioning of graphs. *IBM J Res Dev.* 1973;17(5):420-425.
46. Daoutidis P, Tang W, Allman A. Decomposition of control and optimization problems by network structure: Concepts, methods, and inspirations from biology. *AIChE J.* 2019;65(10):e16708.
47. Haney F, Donovan G, Roth T, Lowrie A, Morlidge G, Lucchini S, Halvorsen S. Modular Processing Facility
48. Bhattacharyya D, Shaeiwitz JA, Whiting WB, Bailie RC, Turton R. *Analysis, Synthesis, and Design of Chemical Processes*. Upper Saddle River, USA: Prentice Hall; 2012.
49. Daoutidis P, Allman A, Khatib S, et al. Distributed decision making for intensified process systems. *Curr Opin Chem Eng.* 2019;25:75-81.
50. Federal Size Regulations for Commercial Motor Vehicles. https://ops.fhwa.dot.gov/freight/publications/size_regs_final_rpt/index.htm. 2017. Accessed July 5, 2019.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Shao Y, Zavala VM. Modularity measures: Concepts, computation, and applications to manufacturing systems. *AIChE J.* 2020;e16965. <https://doi.org/10.1002/aic.16965>