Lessons from social network analysis to Industry 4.0

Yamila M. Omar*, Meysam Minoufekr, Peter Plapper

Faculty of Science, Technology and Communication, University of Luxembourg, Luxembourg

Abstract

With the advent of Industry 4.0, a growing number of sensors within modern production lines generate high volumes of data. This data can be used to optimize the manufacturing industry in terms of complex network topology metrics commonly used in the analysis of social and communication networks. In this work, several such metrics are presented along with their appropriate interpretation in the field of manufacturing. Furthermore, the assumptions under which such metrics are defined are assessed in order to determine their suitability. Finally, their potential application to identify performance limiting resources, allocate maintenance resources and guarantee quality assurance are discussed.

Keywords: complex networks, smart manufacturing, Industry 4.0

2017 MSC: 00-01, 99-00

1. Introduction

Manufacturing systems have evolved from in-series production lines comprised of ordered, sequential, task-specific workstations, towards manufacturing networks made of flexible value-adding units capable of adapting to multiple tasks distinctive of Industry 4.0 [1]. In addition, the automation of repetitive tasks undertaken during the third industrial revolution has been coupled with ubiquitous cyber-physical systems with an ever growing number of embedded sensors that continuously generate high volumes of data [2]. This data is used

*Corresponding author

Email address: `yamila.omar@uni.lu` (Yamila M. Omar)

to optimize manufacturing processes [3] by means of statistical and quantitative analysis, explanatory and predictive modeling, and fact-based decision making known as business analytics [4]. One seemingly unexploited use of such data is the analysis of manufacturing networks by means of complex network topology metrics (CNTM) popular in the study of social and communication networks [5]. Such metrics provide valuable information about individual elements of the network, as well as how they relate to others. In terms of manufacturing, this means that they have the potential to unequivocally identify process limiting resources (or bottlenecks), to aid efficient maintenance resource allocation and to improve quality assurance. In this work, a number of CNTM are presented and their appropriate interpretation in the field of manufacturing networks is proposed. Furthermore, the importance of correctly assessing the assumptions under which such metrics are defined is highlighted, in order to properly interpret results. Finally, potential application areas are suggested where these metrics can aid manufacturing design and optimization.

2. Definitions

Figure 1 (a) shows an illustration of a manufacturing floor plan. As explained earlier, the manufacturing process is traditionally regarded as sequential and therefore, abstracted as process flow charts (see Figure 1, b). However, manufacturing can also be viewed as a complex network. Complex networks are represented as graphs G composed of a set of nodes V and edges E. In the case of manufacturing systems, the nodes represent distinctive workstations while the edges indicate the material flow across them (see Figure 1, c). Since the material flow follows a predetermined path, the edges are said to be directed. A directed graph G can be completely described by its adjacency matrix $\mathscr{A}$, a $N \times N$ matrix, where N is the number of nodes. An entry $a_{ij} = 1$ if there is a link from node i to node j, and zero otherwise [6]. Since modern manufacturing networks are employed in the production of multiple products or even product families, each requiring a different number of workstations and following dis-

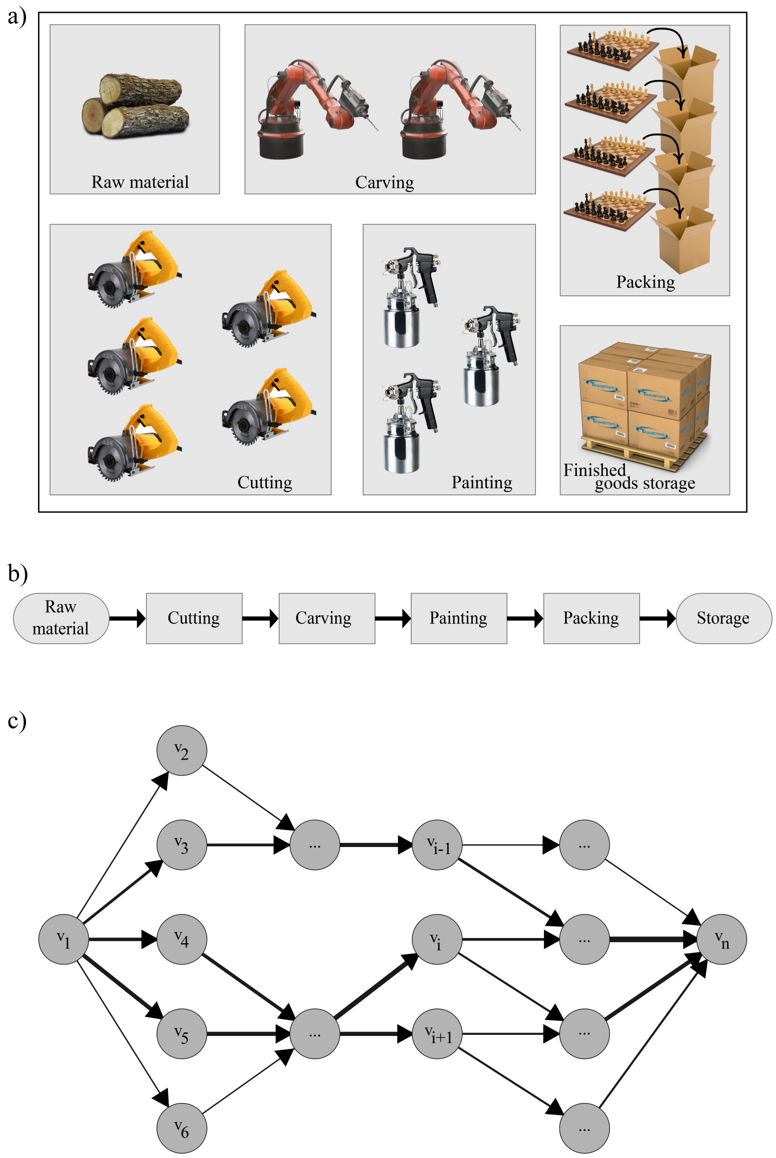

Figure 1: *a)* Manufacturing floor plan showing the location of the raw materials storage room; the cutting, carving, and painting benches; the packing line and the finished goods storage room. *b)* Manufacturing process flow chart showing the sequence of value-adding tasks necessary to transform raw materials into finished goods. *c)* Manufacturing network abstracted as a directed graph G where nodes represent workstations and edges indicate material flows across them. The thickness of the edges, commonly known as their "weight", is proportional to the material flow as indicated by the weight matrix W.

tinct paths through the network, manufacturing systems are better represented as weighted networks [5] characterized by a weight matrix W where each element $0 \leq w_{ij} \leq 1$ indicates the fraction of the total items manufactured that flow from workstation i to j.

3. Complex network metrics

In this section, we present a number of CNTM commonly used in the analysis of social and communication networks [5] and propose the appropriate interpretation when applied to manufacturing networks.

3.1. Node degree

In directed graphs, the in- and out-degree can be defined. The in-degree is the number of ingoing links k_i^{in}, and indicates the number of upstream workstations that i is directly connected to. The out-degree k_i^{out} is the number of outgoing links, and specifies the number of downstream workstations that i is directly linked to. In general, the degree k_i can be calculated as the sum between the in- and out-degree:

$$k_i = k_i^{in} + k_i^{out} = \sum_j a_{ji} + \sum_j a_{ij} \tag{1}$$

Authors analyzing supply chain networks, where companies were represented as nodes and material flows as edges [7], interpreted the in-degree as the degree of difficulty faced by each company when managing incoming flows, i.e. as a metric of each node's operational load coming from upstream suppliers. Likewise, the out-degree was interpreted as the difficulty faced by each node in managing the needs of customer nodes. However, it must be noted that the degree disregards the actual amount of material flow between adjacent nodes, i.e. all edges are considered equally when computing the in- and out-degree. Therefore, in most manufacturing networks, the degree is better regarded as the number of direct neighbors of a given workstation.

3.2. Node strength

In cases where there is a highly heterogeneous material flow between different sets of nodes, the node strength is a more accurate metric of a node's workload. In fact, the node strength is the natural generalization of the degree for weighted graphs [5]. It is defined as

$$s_i = s_i^{in} + s_i^{out} = \sum_j w_{ji} + \sum_j w_{ij} \tag{2}$$

The in- and out-strength represent the supply and demand load of workstation i. Thus, this metric is interpreted as the workload handled by each workstation.

3.3. Betweenness centrality

The betweenness centrality[1] C_B was first defined in [8] as the fraction of times in which a node v falls on the geodesic (shortest) path σ between any two other nodes s and t.

$$C_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \tag{3}$$

Originally introduced to quantify the importance of an individual in a communication network in terms of controlling information flows [8], in the context of manufacturing networks this metric indicates the centrality of a node and its potential to impede or facilitate materials flow through the network [9]. Workstations determined to be structurally central stand between others and therefore exert a high degree of control on the materials flow. It must be noted, however, that the betweenness centrality is calculated under the assumption that nodes of higher importance are located on shortest paths through the network. Such strong assumption is not likely to hold on manufacturing networks and therefore, a different importance measure that forgoes said assumption is introduced in Section 3.5.

[1] The calculation of the betweenness centrality is not trivial. Although a matrix based calculation is described in [6], a faster algorithm was presented by Brandes [10] and later

3.4. Clustering coefficient

The clustering coefficient, firstly introduced in [12], indicates the likelihood that two neighbors of a node i are adjacent, i.e. the ratio between the number of triangles t_i with i as one vertex and the number of all possible triangles that i could form T_i. The original formulation [12] is applicable in the case of binary undirected networks. Several generalizations were made to extend its application to weighted undirected networks [13] as well as to both binary and weighted directed networks [14]. The latter is defined as:

$$\tilde{C}_i^D(A) = \frac{\tilde{t}_i^D}{T_i^D} = \frac{\sum_j \sum_h (w_{ij}^{1/3} + w_{ji}^{1/3})(w_{ih}^{1/3} + w_{hi}^{1/3})(w_{jh}^{1/3} + w_{hj}^{1/3})}{2[k_i(k_i - 1) - 2k_i^{\leftrightarrow}]} \tag{4}$$

where $k_i = k_i^{in} + k_i^{out}$ (see section 3.1) and $k_i^{\leftrightarrow} = \sum_{j \neq i} a_{ij} a_{ji}$ is the number of bilateral edges between i and its neighbors (i.e. the number of nodes j for which both edges, $i \to j$ and $j \to i$, exist). The clustering coefficient of graph G can be easily determined as the average among all nodes in the network, $\tilde{C}^D = N^{-1} \sum_i \tilde{C}_i^D$.

The clustering coefficient describes the type of manufacturing network under study [9]. High values indicate highly interconnected workstations typical of cellular manufacturing, while low values are characteristic of rather serial manufacturing plants.

3.5. PageRank

The PageRank algorithm was originally created to index the World Wide Web [15, 16], which is represented by a complex network of hyperlinks. This iterative calculation converges to the probability distribution $\mathbf{v}'$ of a random walker for all nodes. The most commonly used representation of the PageRank algorithm is that accounting for taxation:

---

extended to the case of weighted networks [11]. It should be noted that "Algorithm 11" presented in [11] for weighted networks contains an error. The accumulation part is missing. A factor of $w(v, w)$ should be applied to $\sigma[v]/\sigma[w]$. The erratum is available in http://algo.uni-konstanz.de/members/brandes/publications/.

$$\mathbf{v}' = \beta M\mathbf{v} + (1-\beta)\mathbf{e}/n \tag{5}$$

where $\mathbf{v}'$ and $\mathbf{v}$ are the probability distribution vectors at the new and previous steps. M is a transition matrix of m_{ij} elements whose values are $1/k$ if node j has k outgoing edges and one points to node i; and zero otherwise. β is a chosen constant (usually in the range between 0.8 and 0.9) that accounts for the random walkers finite probability of leaving the network, $\mathbf{e}$ is a vector of all 1s (i.e. $\mathbf{e}^T = [1, 1, ..., 1]$), and n is the number of nodes. The first term of the equation $\beta M\mathbf{v}$ represents the probability that the walker follows an outgoing edge from the current node, while the second term $(1-\beta)\mathbf{e}/n$ represents the finite probability $(1-\beta)$ of a random walker "jumping" to any other node.[2]

It is noteworthy that while the betweenness centrality (Section 3.3) assumes that important nodes are located on shortest paths through the network, the PageRank algorithm uses a probabilistic approach to determine the likely location of a random walker after one step. The former metric clearly identifies which nodes control information in communication (or social) networks [8], but the PageRank algorithm seems more appropriate when dealing with a manufacturing line where items seldomly follow a geodesic path from start to finish. In fact, the PageRank algorithm considers all permitted steps that a random walker on node i could take with their associated probability as stated in the transition matrix M, to determine the node importance. In other words, the node importance measures the workload build-up of a node while accounting for inter-dependencies among pairs of nodes and thus, can be used to determine bottlenecks.

[2]This second term is known as "taxation" and mathematically handles dead ends (nodes with no outgoing links) and spider traps (nodes with no outgoing links other to themselves). It does not mean that manufactured goods jump between random workstations. A full explanation on this mathematical artifact is available in Chapter 5 of [17].

Table 1: List of useful complex network metrics and their interpretation in terms of manufacturing networks.

Metric	Interpretation
Degree	Number of incoming and outgoing links directly connected to a specific workstation. Based on adjacency matrix, disregards actual material flows.
Strength	Measure of the workload of a node. Based on weight matrix, accounts for material flows.
Betweenness	Centrality measure based on shortest paths between pairs of nodes. Does not consider effective manufacturing paths.
Clustering	Coefficient that measures the degree to which the manufacturing network is interconnected.
PageRank	Probabilistic method that ranks nodes by importance based on effective processing paths (as opposed to shortest paths).

4. Applications

One major use of CNTM in Industry 4.0 is the identification of performance limiting resources [18] usually referred to as bottlenecks. A common approach used to determine bottlenecks is by selecting the workstation with the highest utilization, which is equivalent to identifying the node of highest strength. However, this forgoes the interaction between different workstations. The PageRank metric, for example, which accounts for direct dependencies among nodes, could be used when determining bottlenecks during the design phase of a manufacturing system facilitating design improvements before incurring in capital investments.

Other areas suitable for the application of complex network analysis are maintenance resource allocation and quality assurance. In both cases, determining which workstations are central in the network and which are more likely to affect downstream customers is crucial. This information may help to prevent costly unplanned downtime and propagation of defects along the network.

Centrality metrics, such as the degree, give information about the number of up- and downstream resources a node is connected to; while other metrics, such as the betweenness centrality or the PageRank importance value, show the leverage that each node have on others.

5. Conclusions

Complex networks analysis has a lot to offer to the manufacturing industry. Given current manufacturing data availability, the potential to apply such insights to production network design and optimization is clear. The CNTM presented in this work and the analysis of their underlying assumptions, point to potential applications in identification of bottlenecks, as well as maintenance resource allocation and quality assurance. Further research will certainly generate consensus with respect to the interpretation and application of the various CNTM in modern manufacturing plants.

References

[1] R. Drath, A. Horch, Industrie 4.0: Hit or hype?, IEEE industrial electronics magazine 8 (2) (2014) 56–58. `doi:10.1109/MIE.2014.2312079`.

[2] J. Lee, B. Bagheri, H.-A. Kao, A cyber-physical systems architecture for industry 4.0-based manufacturing systems, Manufacturing Letters 3 (2015) 18–23. `doi:10.1016/j.mfglet.2014.12.001`.

[3] A. Kusiak, Smart manufacturing must embrace big data, Nature 544 (2017) 23–25. `doi:10.1038/544023a`.

[4] T. H. Davenport, J. G. Harris, Competing on analytics: The new science of winning, Harvard Business Press, 2007.

[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Physics reports 424 (4) (2006) 175–308. `doi:10.1016/j.physrep.2005.10.009`.

[6] F. Harary, R. Z. Norman, D. Cartwright, Structural models: An introduction to the theory of directed graphs, John Wiley & Sons Inc, 1965.

[7] Y. Kim, T. Y. Choi, T. Yan, K. Dooley, Structural investigation of supply networks: A social network analysis approach, Journal of Operations Management 29 (3) (2011) 194–211. doi:10.1016/j.jom.2010.11.001.

[8] L. C. Freeman, A set of measures of centrality based on betweenness, Sociometry (1977) 35–41doi:10.2307/3033543.

[9] T. Becker, M. Meyer, K. Windt, A manufacturing systems network model for the evaluation of complex manufacturing systems, International Journal of Productivity and Performance Management 63 (3) (2014) 324–340. doi:10.1108/IJPPM-03-2013-0047.

[10] U. Brandes, A faster algorithm for betweenness centrality, Journal of Mathematical Sociology 25 (2) (2001) 163–177. doi:10.1080/0022250X.2001.9990249.

[11] U. Brandes, On variants of shortest-path betweenness centrality and their generic computation, Social Networks 30 (2) (2008) 136–145. doi:10.1016/j.socnet.2007.11.001.

[12] D. J. Watts, S. H. Strogatz, Collective dynamics of small-world networks, Nature 393 (6684) (1998) 440–442. doi:10.1038/30918.

[13] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, J. Kertesz, Generalizations of the clustering coefficient to weighted complex networks, Physical Review E 75 (2) (2007) 027105. doi:10.1103/PhysRevE.75.027105.

[14] G. Fagiolo, Clustering in complex directed networks, Physical Review E 76 (2) (2007) 026107. doi:10.1103/PhysRevE.76.026107.

[15] S. Brin, L. Page, Anatomy of a large-scale hypertextual web search engine, in: Proc. 7th Intl. World-Wide-Web Conference, 1998, pp. 107–117.

[16] S. Brin, L. Page, Reprint of: The anatomy of a large-scale hypertextual web search engine, Computer Networks 56 (18) (2012) 3825–3833. doi:10.1016/j.comnet.2012.10.007.

[17] J. Leskovec, A. Rajaraman, J. D. Ullman, Mining of massive datasets, Cambridge university press, 2014.

[18] H. Blunck, V. Vican, T. Becker, K. Windt, Improvement heuristics for manufacturing system design using complex network figures, Procedia CIRP 17 (2014) 50–55. doi:10.1016/j.procir.2014.01.063.