Risk Evaluation for a Manufacturing Process Based on a Directed Weighted Network

Lixiang Wang [1] [iD]**, Wei Dai** [1,*] [iD]**, Dongmei Sun** [1,2] **and Yu Zhao** [1]

1   School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China;
    wlx29@buaa.edu.cn (L.W.); sdmei208@163.com (D.S.); zhaoyu@buaa.edu.cn (Y.Z.)
2   No. 208 Research Institute of China Ordnance Industries, Beijing 100191, China
*   Correspondence: dw@buaa.edu.cn; Tel.: +86-010-8233-8673

Abstract: The quality of a manufacturing process can be represented by the complex coupling relationship between quality characteristics, which is defined by the directed weighted network to evaluate the risk of the manufacturing process. A multistage manufacturing process model is established to extract the quality information, and the quality characteristics of each process are mapped to nodes of the network. The mixed embedded partial conditional mutual information (PMIME) is used to analyze the causal effect between quality characteristics, wherein the causal relationships are mapped as the directed edges, while the magnitudes of the causal effects are defined as the weight of edges. The node centrality is measured based on information entropy theory, and the influence of a node is divided into two parts, which are local and indirect effects. Moreover, the entropy value of the directed weighted network is determined according to the weighted average of the centrality of the nodes, and this value is defined as the risk of the manufacturing process. Finally, the method is verified through a public dataset.

Keywords: directed weighted network; PMIME; information entropy; risk evaluation

1. Introduction

In a product manufacturing process, due to the random fluctuations of process factors, deviations in quality characteristics occur. Moreover, random fluctuations are inherent in all stages of the manufacturing process and are difficult to avoid. Therefore, quality control of the manufacturing process has always been an active and important topic [1]. During the manufacturing process, there is a complex coupling relationship between quality characteristics. In general, machining errors for one key characteristic may cause some errors for other characteristics coupled with the key characteristic. Therefore, identifying key quality characteristics is particularly important for quality control in the manufacturing process. Amiri et al. [2] believed that the quality of a manufacturing process is represented by two or more correlated quality characteristics. Obviously, analyzing the relationship between quality characteristics in isolation may lead to erroneous results.

Many scholars have studied the relations between quality characteristics for product design. Quality Function Deployment (QFD) is a common theory and method for product design management driven by customer requirements [3]. The QFD is a process and set of tools used to effectively define customer requirements and convert them into quality characteristics. However, the relations in the design phase are static, this is because the analysis of the quality characteristics of the design process is performed before the product is manufactured. During the manufacturing process, quality characteristics are constantly changing due to changes in many process factors. Therefore, it is not enough to analyze the relationship of quality characteristics from a static state. At the same time, the constant changes in quality characteristics have also brought uncertainty to the risk evaluation of the

manufacturing process. Some quality control methods are presented for the multistage manufacturing process, such as the quality state space theory [4] and stream of variation theory [5]. These methods analyze the change of quality characteristics with the process from the perspective of a time series. It is known that the accumulation of deviations in quality characteristics leads to defects, which are then exposed in the form of early failures during use. Moreover, the relation between quality characteristics is more precisely a causal relationship, because the accumulation of quality characteristic deviations in the previous stage will propagate during the manufacturing process, thereby affecting the quality characteristics in the following stage.

In this paper, the relation of quality characteristics is considered as a directed weighted complex network. The quality characteristics of each process are mapped to nodes of the network. And the partial conditional mutual information (PMIME) is used to analyze the causal effect between quality characteristics of the manufacturing process. Further, the causal relationship between quality characteristics are mapped as a directed edge, while the magnitudes of the causal effects are defined as the weight of edges, thereby a directed weighted network is established. Moreover, the centrality of a node is measured based on information entropy theory. The influence of a node is divided into two parts, which are local and indirect effects. The larger the value of entropy of a node, the greater its influence. Moreover, the entropy $H(I)$ of a single discrete random variable I is a measure of its average uncertainty. For the set of quality characteristics, the entropy of each quality characteristic represents the uncertainty of whether it can complete the requirements of the manufacturing process. That is, the centrality of each node indicates the uncertainty of whether the quality characteristics can meet the requirements of the manufacturing process in a certain time series. In addition, the quality of the manufacturing process actually refers to the degree to which a set of quality characteristics meets its production needs. Therefore, the risk of the manufacturing process is defined as the quality loss caused by the quality characteristics not meeting the production requirements. In this paper, the risk is evaluated by quantifying the uncertainty of the manufacturing process. When the uncertainty of the manufacturing process is greater, the more defects in the manufacturing process and the greater the quality loss. Furthermore, the quality loss is invisible, which means that the reliability of the products produced by the manufacturing process is relatively low. Moreover, the quality loss will spread over the connection between quality characteristics, so the connection between quality characteristics is actually the risk propagation path of the manufacturing process. Therefore, the entropy value of the directed weighted network is determined according to the weighted average of the centrality of the nodes, and the value is defined as the risk of the manufacturing process. Finally, the method is verified through a public dataset.

The outline of this paper is organized as follows. In Section 2, a literature review related to complex network theory is applied to solve and describe complicated manufacturing problems and causality analyses are presented. Section 3 includes details of the proposed approach. Section 4 contains details about a real case to illustrate and verify the proposed method. Finally, the conclusions drawn and discussion are in Section 5.

2. Literature Review

Due to the random fluctuation of process factors, which makes the quality characteristics uncertain, there is a complex causal relationship between quality characteristics in a time series. Therefore, it is essential to establish a model to describe the behavior of the manufacturing process. Some mathematical models that are proposed by researchers that describe the quality information flow in a multistage manufacturing process, the assembly process [6,7] and machining process [8] are included. Hu [9] presented the stream of variation theory for automotive body assembly. Then the variation flow theory was extended to the machining process [10–12]. However, in the actual manufacturing process, quality characteristics affect each other. Therefore, analysis of the error transfer and accumulation of quality characteristics in a multistage manufacturing process may lead to erroneous results. Jiao and Djurdjanovic [13] presented the compensability of error in product quality to eliminate quality

errors accumulated in a multistage manufacturing process. These methods provide some opportunities for the development of modeling and quality control in a multistage manufacturing process.

Many researchers have applied complex network theory to describe and solve complicated manufacturing problems. Based on the complexity of multistage processes, some network models have been established to improve product quality. Lin and Chang [14] focused on performance evaluation of a manufacturing system with multiple production lines based on the network-analysis perspective. Wang et al. [15] proposed an extended machining error propagation network model to quantitatively analyze the complex coupling relationship in the small-batch, multistage machining process of aircraft landing gear parts. Liu et al. [16] used a machining error propagation network of multistage machining processes to describe complicated interactions among different stages. Cheng and Chu [17] proposed a network-based assessment approach for change impacts on complex products, and three changeability indices were presented. Qin et al. [18] utilized a weighted network of multistage machining processes to quantitatively analyze variation propagation. Moreover, the variation propagation stability was analyzed by a virus-spreading model. Kim et al. [19] proposed a product network analysis, which explored a network-leveled relation among all products. Wang et al. [20] proposed a novel approach to support failure mode, effects, and criticality analysis for multistage processes based on complex networks. Di Bona et al. [21] proposed a total efficient risk priority number method that integrated the failure mode, effects, and criticality analysis with other important factors in risk assessment. The above complex network-based modeling analyzed the quality evaluation and control of multistage processes from different levels.

Nevertheless, there is a lack of analysis using quality characteristics to describe the quality of multistage processes. Actually, the accumulation of deviations in quality characteristics leads to the occurrence of defects, which are exposed in the form of early failures during the use phase. Moreover, the quality of a product or process is characterized by monitoring correlated profile and multivariate quality characteristics. Therefore, it is important to analyze the behavior of multistage processes using quality characteristics. Du et al. [22] defined the key product characteristics and designed a model for a key characteristics management system. Köksoy [23] presented a method to optimize multiple quality characteristics based on the mean square error criterion. Ouyang et al. [24] used a QCAC-Entropy-TOPSIS approach to measure quality characteristics and rank improvement priorities for all substandard quality characteristics. Li et al. [25] proposed a key quality characteristics selection technique for imbalanced production data using a two-phase, bi-objective feature selection method. Diao et al. [26] analyzed the coupling relations among quality characteristics and proposed a weighted-coupled, network-based quality control method for improving key characteristics in the product manufacturing process.

However, these methods ignore the causal relationship between quality characteristics and the direction of causal effects. This is because the accumulation of deviations in quality characteristics will propagate during the manufacturing process and affect the quality characteristics in the next stage. Generally, Granger causality [27] and transfer entropy [28,29] are two classic methods for causal analysis. However, these two classic causal analysis methods are only suitable for a bivariate time series. With the development of multivariate state space reconstruction, different embedding schemes [30–33] are used in Granger causality and transfer entropy. Multivariate time series embedding includes uniform and non-uniform embedding. Uniform embedding may have problems such as overfitting and false influence detection. A nonuniform embedding scheme solves the above problems perfectly [34]. Vlachos et al. [35] presented a causality measure of conditional mutual information from mixed embedding (MIME) for bivariate time series. Further, Kugiumtzis [36] extended the MIME method to a multivariate time series, which is named as partial MIME (PMIME).

3. Proposed Method

3.1. Preliminaries

In a multistage manufacturing process, there is a complex causal relationship between quality characteristics. Moreover, due to the random fluctuation of process factors, each quality characteristic is uncertain in a time series. Therefore, if we want to study the effect of the causal relationship between quality characteristics on the quality of manufacturing process, we should first model the multistage manufacturing process and extract the corresponding quality characteristics. Suppose that k is the number of machining stages, and the stage set is written as $S = \{S_1, S_2, \ldots, S_k\}$. Moreover, $\varepsilon = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k\}$ and $\tau = \{\tau_1, \tau_2, \ldots, \tau_k\}$ are defined as the number of individual quality characteristics extracted for a single machining stage and the length of time for each machining stage. Hence, a model of a multistage manufacturing process is shown in Figure 1.
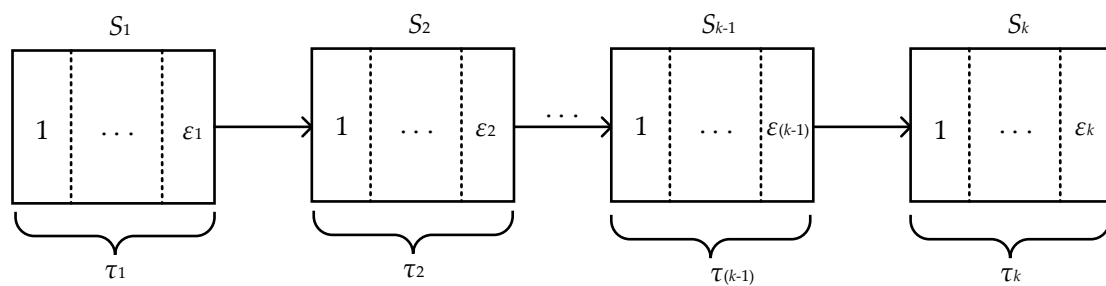

Figure 1. A model of a multistage manufacturing process.

Suppose that n is the number of types of quality characteristics, and the characteristic set is written as $= \{C_1, C_2, \ldots, C_n\}$. Therefore, quality characteristics are extracted from the multistage manufacturing process. This is shown in Figure 2.
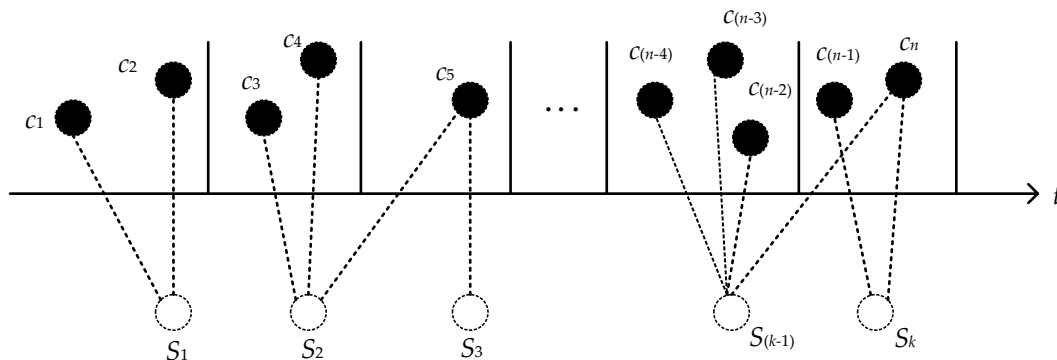

Figure 2. Extracting quality characteristics from a multistage manufacturing process.

Therefore, a directed weighted network is represented as following.

$$G = (N, E, W) \tag{1}$$

where $N = \{N_1, N_2, \ldots, N_n\}$, and N denotes the set of nodes in a given directed weighted network, $E = \{e_{11}, e_{12}, \ldots, e_{ij}, \ldots\}_{n \times n} (1 \leq i \leq n, 1 \leq j \leq n)$, and E represents the set of directed edges from one node to another, $W = \{w_{11}, w_{12}, \ldots, w_{ij}, \ldots\}_{n \times n} (1 \leq i \leq n, 1 \leq j \leq n)$, and set W corresponds to the weighted values. In this paper, nodes are denoted as the quality characteristics of a multistage manufacturing process. Obviously, there is a one-to-one correspondence between the characteristic set C and the node set N. The edges set represents causality among quality characteristics $i \rightarrow j$, and set W of weighted values indicates the magnitude of the causal effect sent from quality characteristic $i \rightarrow j$. When $w_{ij} > 0$, which means that node i has a causal effect on node j, and also means edge e_{ij} exists.

Otherwise $w_{ij}= 0$, which means that node *i* has no causal effect on node *j*, and also means edge e_{ij} does not exist. We assume a set of quality characteristics for a multistage manufacturing process, where $C = \{ C_1, C_2, C_3, C_4, C_5 \}$. A simple example of a directed weighted network is shown in Figure 3.
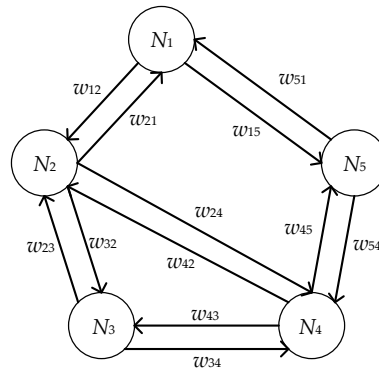

Figure 3. A simple example of a directed weighted network.

3.2. PMIPE Method

In this paper, the PMIME method is used to estimate the direct and directional coupling in a multivariate time series. Let $\{x_t, y_t, z_{1,t}, \dots, z_{n-2,t}\}_{t=1}^{\sum_{i=1}^{k} \tau_i}$ be a multivariate time series of *n* variables $X, Y, Z_1, Z_2, \dots, Z_{n-2}$. What we intend to estimate is the casual effect of *X* on *Y* conditioning on $Z = \{ Z_1, Z_2, \dots, Z_{n-2}\}$. Hence, the driving subsystem is *X* and the response subsystem is *Y*, and *Z* is defined as the other subsystem. Obviously, subsystems *X, Y, Z* contain all quality characteristics of the multistage manufacturing process.

Moreover, the future of variable *X* is represented by a vector of *T* future values, that is $x_t^T= [x_{t+1}, x_{t+2}, \dots, x_{t+T}], 1 \le T \le \sum_{i=1}^{k} \varepsilon_i$. In addition, the lags of *X, Y* and *Z* are sought in a range given by a maximum lag for each variable, e.g., L_x for *X* and L_y for *Y*. Generally, the maximum lag *L* of all variables is equal, that is $L_x= L_y= L_z$. V_t is indicated as the set of all lagged variables at time *t*, the part of $x_t, x_{t-1}, \dots, x_{t-L_x}$ of *X* and the same for subsystem *Y* and *Z* are contained in V_t. An iterative technique is performed to form the mixed embedding vector $v_t \in V_t$. The detailed method is shown in the following steps.

Step 1: starting with an empty embedding vector $v_t^0 = \varnothing$.

Step 2: in the first iteration in order to find the most related to y_t^T in V_t, the embedding vector is represented as v_t^1, which is written as follows:

$$v_t^1 = \underset{v \in V_t}{argmax} I(y_t^T ; v) \tag{2}$$

where $I(.)$ denotes mutual information (MI). And MI is estimated by the k-nearest neighbors (k-NNs) method. Then $w_t^1= [v_t^1]$ is obtained, simultaneously v_t^1 is removed from V_t.

Step 3: in the *m*-th ($m \ge 2$) embedding cycle, the mixed embedding vector is augmented by the component v_t^m of V_t, giving most information about y_t^T additionally to the information already contained in $w_t^{m-1}= [v_t^1, v_t^2, \dots, v_t^{m-1}]$. As for step 2, v_t^m is denoted as follows:

$$v_t^m = \underset{v \in V_t}{argmax} I(y_t^T ; v|w_t^{m-1}) \tag{3}$$

For example, in the second iteration, v_t^2 is written as follows:

$$v_t^2 = \underset{v \in V_t}{argmax} I(y_t^T ; v|w_t^1) \tag{4}$$

where the conditional mutual information (CMI) is calculated by the k-NNs method, and the mixed embedding vector is $w_t^2 = [v_t^1, v_t^2]$. Iterating occurs according to Equation (3) until the additional information of v_t^m selected at the embedding cycle m is not large enough. Moreover, the termination criterion is quantified as:

$$I(y_t^T; w_t^{k-1}) / I(y_t^T; w_t^k) > A \qquad (5)$$

where the threshold $A < 1$ and the value of A is generally 0.95 or 0.97 in [35,36]. Further, the mixed embedding vector $w_t = w_t^{k-1}$ is obtained.

Step 4: calculating the causality effect of X on Y conditioned on the other variables in Z, the PMIME is described as

$$R_{X \to Y|Z} = I\left(y_t^T; w_t^x | w_t^y, w_t^z\right) / I\left(y_t^T; w_t\right) \qquad (6)$$

where w_t^x is denoted as the part of X in w_t, and the same with w_t^y and w_t^z.

3.3. Entropy-Based Centrality Measurement

Based on the PMIME method, the direction of the causal relation and the magnitude of the causal effect of each quality characteristic can be determined. Moreover, each quality characteristic is mapped to a node, the direction of the causal relationship between nodes is mapped to directed edges, and the magnitude of the causal effect between nodes is mapped to the weight of the edges. Hence, a directed weighted network is defined. Furthermore, Qiao et al. [37] proposed an entropy-based centrality measurement to identify the vital node. The total power of a node is divided into two parts, including its local power and its indirect power. The detailed method is described below as follows:

Step 1: a complete directed weighted network is deconstructed into several subnets centered on certain nodes.

Step 2: calculating structural entropy (SE), the SE takes advantage of topographic properties of the sub-graph, evaluating the strength of a given node in specific subnet. Above all, the subnet degree centrality of node i and its neighbor j, which is indicated as SDC_i. This is written as

$$SDC_i = DC_i^{in} + DC_i^{out} \qquad (7)$$

where DC_i^{in} denotes the in-degree centrality of node i (the number of nodes pointing to node i) and DC_i^{out} represents the out-degree centrality of node i (the number of directed edges from node i to another node). Moreover, the SE for node i in subnet G_i is indicated as follows:

$$SE_i = -\sum_{i=1}^{M+1} \frac{SDC_i}{\sum_{i=1}^{M+1} SDC_i} \log \frac{SDC_i}{\sum_{i=1}^{M+1} SDC_i} \qquad (8)$$

where M refers to the number of nodes directly connected to node i in subnet G_i.

Step 3: calculating frequency entropy (FE), the FE takes advantage of the weighted edges that reflects the interaction frequency between two nodes. Further, the FE for node i in subnet G_i is stated as follows:

$$FE_i = -\sum_{j=1}^{H} \frac{W_{ij}}{\sum_{k=1}^{H} W_{ik}} \log \frac{W_{ij}}{\sum_{k=1}^{H} W_{ik}} \qquad (9)$$

where W_{ij} indicates the weight of a directed edge in the given direction and H is the number of node i points to other nodes.

Step 4: Combining Equations (8) and (9), the local power of node i is denoted, which is named as LE_i. Moreover, the LE_i is stated as follows:

$$LE_i = \theta_1 SE_i + \theta_2 FE_i \qquad (10)$$

where θ_1 and θ_2 are the weight coefficients respectively, and $\theta_1 + \theta_2 = 1$.

Step 5: Calculating the indirect power of node *i* on its second-order neighbor node *k*, which is denoted as IE_{ik}. This is written as follows:

$$IE_{ik} = -\sum_{k=1}^{N_{ik}} \frac{LE_i \times LE_s}{N_{ik}} \tag{11}$$

where N_{ik} is the number of first-order neighbor nodes between node *i* and *k*. And LE_s represents the local power of node *s*, and node *s* connects nodes *i* and *k*. We take a two-path subnet with a quadrilateral structure as an example, which is shown in Figure 4.
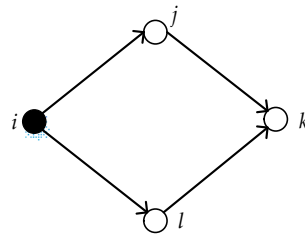

Figure 4. A two-path subnet.

As described above, node *s* represents node *j* and *l* in Figure 4. Moreover, the indirect power of node *i* on its second-order neighbor node *k* is denoted as follows:

$$IE_{ik} = \frac{LE_i \times LE_j + LE_i \times LE_l}{2} \tag{12}$$

Hence, the total indirect power of node *i* on its second-order neighbor nodes is denoted as IE_i, which is stated as follows:

$$IE_i = \frac{\sum_{k=1}^{H_i} IE_{ik}}{H_i} \tag{13}$$

where H_i is the total number of second-order neighbor nodes of node *i*.

Step 6: in line with Equations (10) and (13), the total power of node *i* is represented by E_i, which is named as follows:

$$E_i = \mu_1 LE_i + \mu_2 IE_i \tag{14}$$

where μ_1 and μ_2 are the weight coefficients respectively, and $\mu_1 + \mu_2 = 1$.

3.4. Risk Evaluation

Burduk and Chlebus [38] thought of risk as the danger of failing to achieve the goals specified in the decision. In a multistage manufacturing process, the quality can be denoted as the sum of the characteristics of the process capability to meet explicit and implicit needs. Moreover, entropy is a measure of the uncertainty of the state of quality characteristics in the manufacturing process, that is, a measure of quality loss. Hence, risk is defined as the quality entropy of a multistage manufacturing process. Moreover, since the multistage manufacturing process can be represented by a directed weighted network, the risk of the multistage manufacturing process is defined as the weighted average of the centrality of the nodes. Further, the node weight is calculated as follows:

$$\lambda_i = \frac{SDC_i}{\sum_{i=1}^{N} SDC_i} \tag{15}$$

where, SDC_i represents the degree of node *i* and *N* is the number of nodes. In addition, $\sum_{i=1}^{N} \lambda_i = 1$. Thus, the risk of a multistage manufacturing process is written as follows:

$$R = \sum_{i=1}^{N} \lambda_i E_i \tag{16}$$

4. Case Study

The data of the case come from the SECOM dataset of the UCI Machine Learning Repository [39], which is about a semi-conductor manufacturing process and has 1567 samples, each sample with 591 quality characteristics. The first 59 quality characteristics of the data set were extracted as an example to illustrate the algorithm in this paper. In other words, $\sum_{i=1}^{k} \tau_i = 1567$ and $n = 59$. Moreover, causality between 59 quality characteristics was determined based on PMIME, and we used $A = 0.95$, $L = 5$ and $T = 3$. Further, a directed weighted network was setup, which is shown in Figure 5.
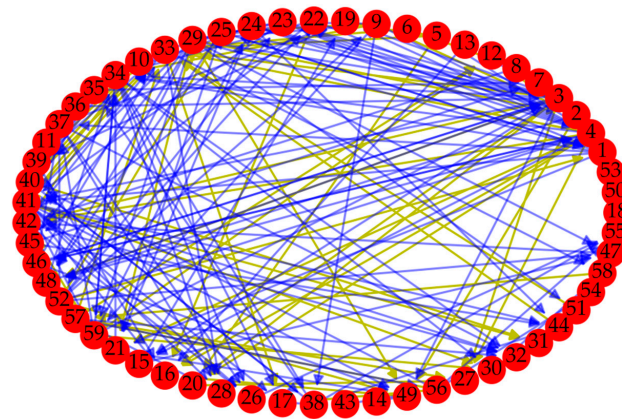

Figure 5. The directed weighted network of case.

Where the yellow line indicates that the weight of the edge is greater than 0.5, and the blue line indicates that the weight of the edge is less than or equal to 0.5. Moreover, nodes 18, 50, and 53 are isolated and not connected to other nodes. In addition, the degree distribution of nodes is shown in Figure 6.
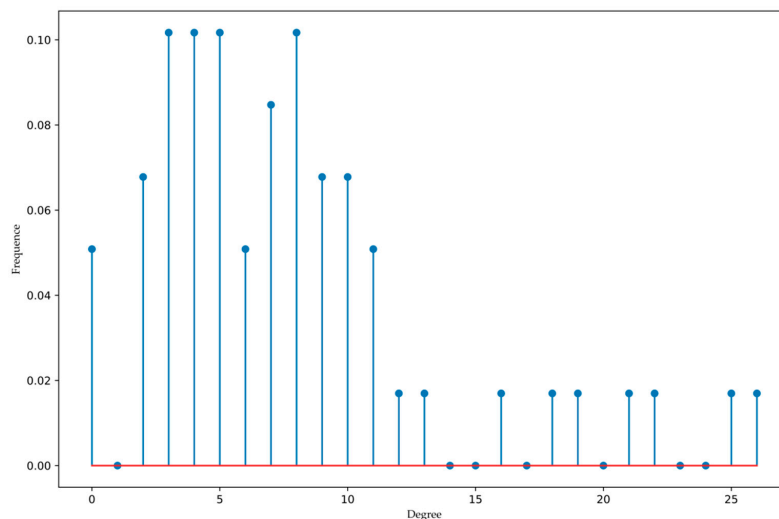

Figure 6. The degree distribution of nodes.

According to Figure 6, it can be concluded that the degree distribution of nodes does not have much regularity, which is neither like Poisson distribution nor power-law distribution. More samples may be needed to further observe the statistical characteristics of the degree distribution. Moreover, only a few nodes have a higher degree. We took the subnet of node 1 as an example to explain the calculating process of the proposed algorithm, and the results of SDC_i as shown in Table 1.

Table 1. The results of SDC_i of nodes in subnet G_1.

Node	DC_i^{in}	DC_i^{out}	SDC_i
1	4	6	10
4	2	3	5
21	5	1	6
23	4	2	6
25	4	0	4
29	3	0	3
34	4	1	5
37	3	4	7
41	1	7	8
52	1	7	8

Based on Equation (8), 10 is set as the base of the logarithmic function, then the structural entropy of node 1 is described as follows:

$$SE_1 = -\sum_{i=1}^{10} \frac{SDC_i}{\sum_{i=1}^{10} SDC_i} log \frac{SDC_i}{\sum_{i=1}^{10} SDC_i} = 0.9774 \tag{17}$$

Moreover, the frequency entropy is calculated by Equation (9), which is written as follows:

$$FE_1 = -\sum_{j=1}^{6} \frac{W_{ij}}{\sum_{k=1}^{6} W_{ik}} log \frac{W_{ij}}{\sum_{k=1}^{6} W_{ik}} = 0.7108 \tag{18}$$

In Equation (10), the θ_1 and θ_2 are set as 0.4 and 0.6, respectively and the local influence of node 1 is defined as follows:

$$LE_1 = 0.4 \, SE_1 + 0.6 \, FE_1 = 0.8174 \tag{19}$$

Following Equations (11) and (13), the indirect influence of node 1 is stated as follows:

$$IE_1 = \frac{\sum_{k=1}^{30} IE_{ik}}{30} = 0.9242 \tag{20}$$

Further, the μ_1 and μ_2 are denoted as 0.6 and 0.4 particularly. Then the total influence of node 1 is expressed as follows:

$$E_1 = 0.6 \, LE_1 + 0.4 IE_1 = 0.8601 \tag{21}$$

Hence, the total power of each node and the corresponding ranking results are shown in Table 2.

Further, the comparison of the local power, indirect power, and total power of each node is shown in Figure 7.
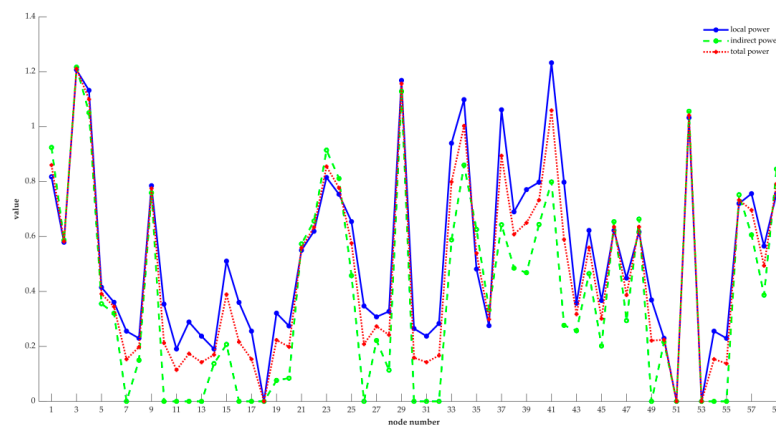

Figure 7. The comparison of the local power, indirect power, and total power of each node.

Table 2. The total power of each node and the corresponding ranking results.

Node	LE_i	IE_i	E_i	No.	Node	LE_i	IE_i	E_i	No.
1	0.8174	0.9242	0.8601	8	31	0.2373	0	0.1424	51
2	0.5791	0.5867	0.5821	23	32	0.2833	0	0.1670	48
3	1.2070	1.2165	1.2108	1	33	0.9396	0.5878	0.7989	10
4	1.1320	1.0504	1.0994	3	34	1.0982	0.8598	1.0028	6
5	0.4150	0.3554	0.3912	29	35	0.4815	0.6256	0.5392	27
6	0.3607	0.3206	0.3447	32	36	0.2756	0.3326	0.2984	35
7	0.2555	0	0.1533	50	37	1.0614	0.6429	0.8940	7
8	0.2295	0.1497	0.1976	45	38	0.6898	0.4848	0.6078	21
9	0.7851	0.7593	0.7748	13	39	0.7708	0.4683	0.6498	17
10	0.3538	0	0.2123	42	40	0.7977	0.6437	0.7325	15
11	0.1908	0	0.1145	53	41	1.2324	0.7986	1.0589	4
12	0.2889	0	0.1734	46	42	0.7977	0.2768	0.5893	22
13	0.2373	0	0.1424	51	43	0.3571	0.2571	0.3171	33
14	0.1908	0.1374	0.1694	47	44	0.6222	0.4650	0.5593	25
15	0.5103	0.2072	0.3891	30	45	0.3673	0.2016	0.3010	34
16	0.3601	0	0.2161	41	46	0.6222	0.6539	0.6349	19
17	0.2555	0	0.1533	50	47	0.4486	0.2941	0.3868	31
18	0	0	0	54	48	0.6166	0.6633	0.6353	18
19	0.3210	0.0762	0.2231	39	49	0.3689	0	0.2213	40
20	0.2746	0.0844	0.1985	44	50	0	0	0	54
21	0.5498	0.5730	0.5591	26	51	0.2295	0.2156	0.2240	38
22	0.6191	0.6564	0.6340	20	52	1.0325	1.0557	1.0412	5
23	0.8151	0.9143	0.8548	9	53	0	0	0	54
24	0.7538	0.8111	0.7767	12	54	0.2555	0	0.1533	50
25	0.6542	0.4576	0.5756	24	55	0.2295	0	0.1377	52
26	0.3470	0	0.2082	43	56	0.7201	0.7518	0.7328	14
27	0.3074	0.2214	0.2730	36	57	0.7560	0.6060	0.6960	16
28	0.3274	0.1136	0.2419	37	58	0.5651	0.3868	0.4938	28
29	1.1683	1.1283	1.1563	2	59	0.7560	0.8454	0.7918	11
30	0.2649	0	0.1589	49					

Combining Table 2 and Figure 7, in the whole network, nodes 3, 4, 29, 34, 41, and 52 are obviously more important than the other nodes. Moreover, the values of entropy-based centrality of these six nodes are all bigger than 1, while the total power of the 7th ranked node is 0.8940, which is a big gap with the top six nodes. Hence, nodes 3, 4, 29, 34, 41, and 52 are considered as vital nodes, for their changes have a greater influence on the nature of the directed weighted network, and they affect more nodes, too. Moreover, the value of entropy-based centrality of 32 nodes is less than 0.5, which is more than half of the data set. And only nine nodes are greater than 0.8, which includes nodes 1, 3, 4, 23, 29, 34, 37, 41 and 52. This is consistent with the information given in Figure 6. That is, only a few nodes have a higher degree and can affect more nodes. Therefore, the risk of the manufacturing process can be reduced by improving the quality of key nodes. Moreover, the risk of the manufacturing process is defined by the weighted average of the centrality of each node. The weight of each node is obtained through the degree of the node. The greater the degree of the node, the higher the weight, and the greater the probability that the node poses a risk to the whole manufacturing process. Further, based on Equation (15), the degree of each node and their weights are shown in Table 3.

Furthermore, following Equation (16), the risk of the whole network is defined as follows:

$$R = \sum_{i=1}^{59} \lambda_i E_i = 0.6390 \qquad (22)$$

In addition, as shown in Table 2, the biggest value of entropy-based centrality of nodes is 1.2108. Hence, the risk scope of the whole manufacturing process is from 0 to 1.2108. Supposing the range is divided into three sets from small to large, which is [0, 0.4036), [0.4036, 0.8072) and [0.8072, 1.2108]. The corresponding risks are low, medium, and high, therefore the risk of manufacturing process is

medium in this case. Moreover, nodes 3, 4, 29, 34, 41, and 52 account for 25.77% of the weight set. Obviously, controlling these six nodes is critical to improving the quality of the manufacturing process.

Table 3. The degree of each node and their weights.

Node	SDC_i	λ_i	Node	SDC_i	λ_i	Node	SDC_i	λ_i
1	10	0.0206	21	8	0.0165	41	25	0.0515
2	7	0.0144	22	7	0.0144	42	11	0.0227
3	26	0.0536	23	12	0.0247	43	2	0.0041
4	21	0.0433	24	8	0.0165	44	7	0.0144
5	3	0.0062	25	9	0.0186	45	6	0.0124
6	2	0.0041	26	7	0.0144	46	10	0.0206
7	4	0.0082	27	5	0.0103	47	5	0.0103
8	3	0.0062	28	6	0.0124	48	10	0.0206
9	9	0.0186	29	18	0.0371	49	8	0.0165
10	7	0.0144	30	42	0.0866	50	0	0.0000
11	2	0.0041	31	3	0.0062	51	3	0.0062
12	5	0.0103	32	5	0.0103	52	13	0.0268
13	3	0.0062	33	4	0.0082	53	0	0.0000
14	2	0.0041	34	22	0.0454	54	4	0.0082
15	5	0.0103	35	9	0.0186	55	3	0.0062
16	8	0.0165	36	4	0.0082	56	9	0.0186
17	4	0.0082	37	16	0.0330	57	8	0.0165
18	0	0.0000	38	8	0.0165	58	5	0.0103
19	6	0.0124	39	10	0.0206	59	11	0.0227
20	4	0.0082	40	11	0.0227			

5. Conclusions and Discussion

A key quality characteristic affects many other quality characteristics in a multistage manufacturing process, and the fluctuation of key quality characteristics in the manufacturing process makes the quality characteristics affected by it deviate. Therefore, analyzing the individual quality characteristics in isolation may cause a large deviation in the risk of the manufacturing process. Moreover, the quality characteristics are constantly changing over time. In this paper, the set of quality characteristics is divided into three subsystems: the driving subsystem, response subsystem, and the other subsystem. And the PMIME is used to mine the causality between quality characteristics in the time series. Further, based on complex networks theory, the causal relationship between quality characteristics is mapped to a directed edge, while individual quality characteristics are mapped to nodes, and the magnitudes of the causal effects are defined as the weight of edges. Then a directed weight is established, and the power of a node is divided into two parts, which are local and indirect effects. An entropy-centrality approach is applied to rank influential nodes. This method innovatively solves the possible problem of determining the edges of the complex network through static qualitative analysis because the status of nodes and the status between nodes are updated at any time. In addition, the quality of the manufacturing process is represented by two or more correlated quality characteristics. Hence, a novel index for evaluating risk based on the entropy-centrality of nodes of the complex network has been proposed. This indicator reflects the risk of the manufacturing process, and its size is constantly updated with the change of the state of the nodes and the magnitude and direction of the edge. Therefore, the risk of the manufacturing process can be better controlled according to the change of indicator.

According to Equation (22), the risk of the manufacturing process is 0.6390. First of all, we should determine the acceptable threshold of risk according to the actual situation. If the risk is acceptable, no risk control measures are required for the manufacturing process. Otherwise, we should take risk control measures for the manufacturing process. Second, nodes 3, 4, 29, 34, 41, and 52 are identified as key nodes, and controlling the risks of these six nodes can effectively reduce the risk of the entire manufacturing process. According to the subnets of these six nodes, other quality characteristics that

are causally related to these six quality characteristics can be quickly determined. We can change the processing technology, etc., so that the causal chain of quality characteristics is interrupted, that is, the propagation path of the risk is cut off. For example, node 1 affects node 21, and node 21 has an effect on other nodes. As long as the impact on node 1 or node 21 is cut off, the risk propagation path from node 1 to node 21 is interrupted, and the risk of node 1 is reduced. The example can be used to verify the effectiveness of our proposed risk control strategy. Hence, the structural entropy of node 1 is described as follows:

$$SE_1 = -\sum_{i=1}^{9} \frac{SDC_i}{\sum_{i=1}^{9} SDC_i} log \frac{SDC_i}{\sum_{i=1}^{9} SDC_i} = 0.9271 \tag{23}$$

And the frequency entropy of node 1 is calculated as follows:

$$FE_1 = -\sum_{j=1}^{5} \frac{W_{ij}}{\sum_{k=1}^{5} W_{ik}} log \frac{W_{ij}}{\sum_{k=1}^{5} W_{ik}} = 0.6235 \tag{24}$$

Hence, the local influence of node 1 is 0.7449. Moreover, the indirect influence of node 1 is stated as follows:

$$IE_1 = \frac{\sum_{k=1}^{29} IE_{ik}}{29} = 0.8806 \tag{25}$$

Further, the total influence of node 1 is 0.7992 and the structural entropy of node 21 is changed as 0.8814. Then the local influence and the indirect influence of node 21 is 0.5290 and 0.5513, respectively. Further, the total influence of node 21 is 0.5379. In addition, the weights of all nodes will also change, which is shown in Table 4.

Table 4. The weights of each node.

Node	λ_i	Node	λ_i	Node	λ_i	Node	λ_i	Node	λ_i
1	0.0187	13	0.0062	25	0.0187	37	0.0333	49	0.0146
2	0.0146	14	0.0042	26	0.0146	38	0.0166	50	0.0000
3	0.0541	15	0.0104	27	0.0104	39	0.0208	51	0.0062
4	0.0437	16	0.0166	28	0.0125	40	0.0229	52	0.0270
5	0.0062	17	0.0083	29	0.0374	41	0.0520	53	0.0000
6	0.0042	18	0.0000	30	0.0873	42	0.0229	54	0.0083
7	0.0083	19	0.0125	31	0.0062	43	0.0042	55	0.0062
8	0.0062	20	0.0083	32	0.0104	44	0.0146	56	0.0187
9	0.0187	21	0.0146	33	0.0083	45	0.0125	57	0.0166
10	0.0146	22	0.0146	34	0.0437	46	0.0208	58	0.0104
11	0.0042	23	0.0249	35	0.0187	47	0.0104	59	0.0229
12	0.0104	24	0.0166	36	0.0083	48	0.0208		

Finally, the risk of the manufacturing process is updated to 0.6375, which is lower than before. Since there are many edges between nodes in this paper, although the risk propagation paths of nodes 1 and 21 are cut off, the role of risk control for the entire manufacturing process is relatively small. At the same time, it also proves that cutting off the risk propagation path of node 1 and node 21 has a practical effect on the risk control of the manufacturing process. Moreover, similar risk control measures are continuously taken for other quality characteristic causal chains until the risk of the manufacturing process reaches an acceptable level.

In this study, PMIME is used to detect the direct causality in quality characteristics, however, the curse of dimensionality, resulting in inaccurate estimates of mutual information as the embedding space increases, is unavoidable. Thus, for future work, we expect to carry out further work on improving the accuracy of the algorithm under the premise of dimensionality reduction. In addition, the algorithm in this paper is suitable for a large sample size, but may not be suitable for a small sample size. Therefore, exploring the establishment of a complex network in the case of small samples and ranking of the importance of nodes is also a question worthy of further investigation. Moreover, identifying how

process factors affect the quality of the manufacturing process is also a topic worth studying in future work. That is, what is the mechanism of the quality characteristic deviation caused by the process factors, which will have a great impact on the quality control? Furthermore, the PMIME method in this paper can be used for causality analysis between process factors and quality characteristics, which will make the research in this paper more significant.

Author Contributions: L.W. conceived and designed the methods; W.D. devised and validated the model; Y.Z. analyzed the data; L.W. and D.S. contributed to the writing and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, J.; Freiheit, T.; Hu, S.J.; Koren, Y. A Quality Prediction Framework for Multistage Machining Processes Driven by an Engineering Model and Variation Propagation Model. *J. Manuf. Sci. Eng.* **2007**, *129*, 1088–1100. [CrossRef]
2. Amiri, A.; Zou, C.; Doroudyan, M.H. Monitoring Correlated Profile and Multivariate Quality Characteristics. *Qual. Reliab. Eng. Int.* **2014**, *30*, 133–142. [CrossRef]
3. Chan, L.K. Ming-Lu Wu Quality function deployment: A literature review. *Eur. J. Oper. Res.* **2002**, *143*, 463–497. [CrossRef]
4. Aoki, M.; Havenner, A. State space modeling of multiple time series. *Econom. Rev.* **1991**, *10*, 1–59. [CrossRef]
5. Abellan-Nebot, J.V.; Liu, J.; Subirón, F.R.; Shi, J. State Space Modeling of Variation Propagation in Multistation Machining Processes Considering Machining-Induced Variations. *J. Manuf. Sci. Eng.* **2012**, *134*. [CrossRef]
6. Mantripragada, R.; Whitney, D.E. Modeling and controlling variation propagation in mechanical assemblies using state transition models. *IEEE Trans. Robot. Autom.* **1999**, *15*, 124–140. [CrossRef]
7. Jin, J.; Shi, J. State Space Modeling of Sheet Metal Assembly for Dimensional Control. *J. Manuf. Sci. Eng.* **1999**, *121*, 756–762. [CrossRef]
8. Shiyu Zhou; Qiang Huang; Jianjun Shi State space modeling of dimensional variation propagation in multistage machining process using differential motion vectors. *IEEE Trans. Robot. Autom.* **2003**, *19*, 296–309. [CrossRef]
9. Hu, S.J. Yoram Koren Stream-of-Variation Theory for Automotive Body Assembly. *CIRP Ann. Manuf. Technol.* **1997**, *46*, 1–6. [CrossRef]
10. Barhak, J.; Djurdjanovic, D.; Spicer, P.; Katz, R. Integration of reconfigurable inspection with stream of variations methodology. *Int. J. Mach. Tools Manuf.* **2005**, *45*, 407–419. [CrossRef]
11. Huang, Q.; Shi, J.; Yuan, J. Part Dimensional Error and Its Propagation Modeling in Multi-Operational Machining Processes. *J. Manuf. Sci. Eng.* **2003**, *125*, 255. [CrossRef]
12. Ming, J.; LI, Y.; Tsung, F. Chart allocation strategy for serial-parallel multistage manufacturing processes. *Iie Trans.* **2010**, *42*, 577–588.
13. Jiao, Y.; Djurdjanovic, D. Compensability of errors in product quality in multistage manufacturing processes. *J. Manuf. Syst.* **2011**, *30*, 204–213. [CrossRef]
14. Lin, Y.-K.; Chang, P.-C. Reliability evaluation for a manufacturing network with multiple production lines. *Comput. Ind. Eng.* **2012**, *63*, 1209–1219. [CrossRef]
15. Wang, Y.; Jiang, P.; Leng, J. An extended machining error propagation network model for small-batch machining process control of aircraft landing gear parts. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **2017**, *231*, 1347–1365. [CrossRef]
16. Liu, D.; Jiang, P. The complexity analysis of a machining error propagation network and its application. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2009**, *223*, 623–640. [CrossRef]
17. Cheng, H.; Chu, X. A network-based assessment approach for change impacts on complex product. *J. Intell. Manuf.* **2012**, *23*, 1419–1431. [CrossRef]

18. Qin, Y.; Zhao, L.; Yao, Y.; Xu, D. Multistage machining processes variation propagation analysis based on machining processes weighted network performance. *Int. J. Adv. Manuf. Technol.* **2011**, *55*, 487–499. [CrossRef]

19. Kim, H.K.; Kim, J.K.; Qiu, Y.C. A product network analysis for extending the market basket analysis. *Expert Syst. Appl.* **2012**, *39*, 7403–7410. [CrossRef]

20. Wang, L.; Dai, W.; Luo, G.; Zhao, Y. A Novel Approach to Support Failure Mode, Effects, and Criticality Analysis Based on Complex Networks. *Entropy* **2019**, *21*, 1230. [CrossRef]

21. Di Bona, G.; Silvestri, A.; Forcina, A.; Petrillo, A. Total efficient risk priority number (TERPN): A new method for risk assessment. *J. Risk Res.* **2018**, *21*, 1384–1408. [CrossRef]

22. Du, W.; Mo, R.; Li, S.; Li, B. Design of Product Key Characteristics Management System. *Adv. Mater. Res.* **2012**, *468–471*, 835–838. [CrossRef]

23. Köksoy, O. A nonlinear programming solution to robust multi-response quality problem. *Appl. Math. Comput.* **2008**, *196*, 603–612. [CrossRef]

24. Ouyang, L.Y.; Chen, K.-S.; Yang, C.-M.; Hsu, C.-H. Using a QCAC–Entropy–TOPSIS approach to measure quality characteristics and rank improvement priorities for all substandard quality characteristics. *Int. J. Prod. Res.* **2014**, *52*, 3110–3124. [CrossRef]

25. Li, A.-D.; He, Z.; Wang, Q.; Zhang, Y. Key quality characteristics selection for imbalanced production data using a two-phase bi-objective feature selection method. *Eur. J. Oper. Res.* **2019**, *274*, 978–989. [CrossRef]

26. Diao, G.; Zhao, L.; Yao, Y. A weighted-coupled network-based quality control method for improving key features in product manufacturing process. *J. Intell. Manuf.* **2016**, *27*, 535–548. [CrossRef]

27. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [CrossRef]

28. Marko, H. The Bidirectional Communication Theory-A Generalization of Information Theory. *IEEE Trans. Commun.* **1973**, *21*, 145–1351. [CrossRef]

29. Schreiber; Thomas Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef]

30. Mao, X.; Shang, P. Transfer entropy between multivariate time series. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *47*, 338–347. [CrossRef]

31. Montalto, A.; Stramaglia, S.; Faes, L.; Tessitore, G.; Prevete, R.; Marinazzo, D. Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality. *Neural Netw.* **2015**, *71*, 159–171. [CrossRef] [PubMed]

32. Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2011**, *83*, 051112. [CrossRef] [PubMed]

33. Zhao, X.; Sun, Y.; Li, X.; Shang, P. Multiscale transfer entropy: Measuring information transfer on multiple time scales. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *62*, 202–212. [CrossRef]

34. Sysoev, I.V.; Ponomarenko, V.I.; Pikovsky, A. Reconstruction of coupling architecture of neural field networks from vector time series. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *57*, 342–351. [CrossRef]

35. Vlachos, I.; Kugiumtzis, D. Nonuniform state-space reconstruction and coupling detection. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **2010**, *82*, 016207. [CrossRef]

36. Kugiumtzis, D. Direct-coupling information measure from nonuniform embedding. *Phys. Rev. E* **2013**, *87*, 062918. [CrossRef]

37. Qiao, T.; Shan, W.; Yu, G.; Liu, C. A Novel Entropy-Based Centrality Approach for Identifying Vital Nodes in Weighted Networks. *Entropy* **2018**, *20*, 261. [CrossRef]

38. Burduk, A.; Chlebus, E. Methods of risk evaluation in manufacturing systems. *Arch. Civ. Mech. Eng.* **2009**, *9*, 17–30. [CrossRef]

39. Dua, D.; Graff, C. UCI Machine Learning Repository: SECOM Data Set. Available online: http://archive.ics.uci.edu/ml/datasets/SECOM (accessed on 9 June 2020).