

Article Novel Prediction Model for Steel Mechanical Properties with MSVR Based on MIC and Complex Network Clustering

Yuchun Wu 🗅, Yifan Yan and Zhimin Lv *



Collaborative Innovation Center of Steel Technology, University of Science and Technology Beijing, Beijing 100083, China; s20191303@xs.ustb.edu.cn (Y.W.); b20170502@xs.ustb.edu.cn (Y.Y.) * Correspondence: lvzhimin@nercar.ustb.edu.cn

Abstract: Traditional mechanical properties prediction models are mostly based on experience and mechanism, which neglect the linear and nonlinear relationships between process parameters. Aiming at the high-dimensional data collected in the complex industrial process of steel production, a new prediction model is proposed. The multidimensional support vector regression (MSVR)based model is combined with the feature selection method, which involves maximum information coefficient (MIC) correlation characterization and complex network clustering. Firstly, MIC is used to measure the correlation between process parameters and mechanical properties, based on which a complex network is constructed and hierarchical clustering is performed. Secondly, we evaluate all parameters and select a representative one for each partition as the input of the subsequent model based on the centrality and influence indicators. Finally, an actual steel production case is used to train the MSVR prediction model. The prediction results show that our proposed framework can capture effective features from the full parameters in terms of higher prediction accuracy and is less time-consuming compared with the Pearson-based subset, full-parameter subset, and empirical subset input. The feature selection method based on MIC can dig out some nonlinear relationships which cannot be found by Pearson coefficient.

Keywords: mechanical properties prediction; high-dimensional data; feature selection; maximum information coefficient; complex network clustering

1. Introduction

The level of steel industry is an important indicator to measure the industrialization of the country. At present, all walks of life have more and more stringent requirements for iron and steel products. The mechanical properties of steel can often mean the difference between a long, efficient life in the most abrasive and wear-intensive applications, and frequent or even catastrophic failure. Understanding these properties is absolutely important because all production activities are ultimately to satisfy the actual quality requirements. To maintain and improve the product quality, energy efficiency, and economic profits, the quality prediction and control based on some mechanical properties are essential and have been investigated quite extensively in recent years [1]. Among numerous indicators, tensile strength, yield strength, and elongation are the most commonly used measurements for product's mechanical property, which are affected by a variety of comprehensive factors [2]. However, the production process of steel products contains complex physical and chemical changes with intricate technological processes, which means that property prediction and control have always been a difficult problem in the metallurgical industry. In the traditional practice, property prediction depends on the experience and destructive test, which are costly, time-consuming, and laborious. If the prediction could consider the relevant process parameters, and accordingly optimize the metal composition and process technology, it can greatly reduce the testing time and improve the production efficiency of iron and steel enterprises. Based on this idea, two main methods for that are the empirical and statistical



Citation: Wu, Y.; Yan, Y.; Lv, Z. Novel Prediction Model for Steel Mechanical Properties with MSVR Based on MIC and Complex Network Clustering. *Metals* **2021**, *11*, 747. https:// doi.org/10.3390/met11050747

Academic Editor: Pedro Prates

Received: 6 April 2021 Accepted: 28 April 2021 Published: 1 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). models respectively, while the prediction accuracy can still be improved. This is primarily because these methods mostly depend on the experience and mechanism and ignore the value of data. The parameters selected are too few to represent the actual situation.

By contrast, data-driven prediction methods do not require deep understanding of mechanism but depend on the collected process data only [3]. With the application of big data platform in the steel industry, the real-time data of the throughout production process can be easily obtained, and the number and dimension of samples increase explosively. High-dimensional data have high value theoretically, but they will greatly increase the complexity of modeling and bring the curse of dimensionality [4]. Owing to the process complexity and intricate variable interactions, the major problem is that the nonlinearity and coupling between variables restrict the choice of prediction models and methods. Therefore, how to extract knowledge from the throughout process data, select effective features from the full-parameter set and ultimately establish a more accurate performance prediction model is our main job.

As a dimensionality reduction method, feature selection aims to select the most representative feature subset from the original data set [5], which mainly involves two steps: feature subset selection and feature subset evaluation. The prevailing approaches to feature selection fall into three categories: (a) filters, (b) wrapper methods, and (c) embedded methods. Filter is a single feature selection process, which is independent of subsequent learners. It usually ranks features in the parameter space to obtain subsets [6]. The performance evaluation of the learners works as the evaluation criteria of the wrapper methods. As a representative, Las Vegas Wrapper (LVW) method uses random strategy to search feature subset, and takes the error of the final classifier as the subset evaluation standard [7,8]. Embedded methods integrate feature selection and training learners, which automatically select features during learner training [9]. Moreover, some dimension reduction methods such as principal component analysis (PCA), singular value decomposition (SVD), linear discriminant analysis (LDA), and ISOMAP algorithm can also be regarded as feature selection methods for special basic data [10-13]. However, such methods do not consider the correlation and redundancy between attributes before and after dimensionality reduction, and the results are lack of interpretability.

In fact, when analyzing the relationship between the high-dimensional variables, a variety of distance and similarity indicators can be used to measure the correlation and redundancy between attributes, such as distance, information gain, mutual information, dependency, consistency, etc. Meanwhile, the higher the correlation between attributes, the stronger the necessity and operability of feature selection. Narayana designed an artificial neural network (ANN) model to correlate the complex relations among composition, temperature, and mechanical properties of steels. The ANN predictions are more accurate with experimental results as compared with the calculated properties of the existing model [14,15]. Some studies improve the performance of feature selection by choosing effective measurement indicators [16,17]. Nevertheless, many indicators such as Pearson coefficient, maximum information compression coefficient, and least squares regression error can only measure the linear relationship between features but not the nonlinear relationship. On the basis of information theory, Reshef proposed the concept of maximum information coefficient (MIC), which can widely measure the linear and nonlinear correlation between features and capture many functional or nonfunctional relationships [18]. Moreover, it is confirmed that MIC can accurately measure the correlation between attributes in large data sets. In addition, an intelligent MIC is presented to quickly approach the optimal value [19].

Clustering methods can be used for feature selection, which divide all the nodes in the network into several discrete subgroups based on the correlation metrics [20]. In the theory of complex network, the reason why an actor has power is because of its relationship with other actors. Therefore, it can be considered that the nodes in one cluster have similar "power" or "importance", and nodes with most centrality can be selected as the representation of each partition. If we regard all process parameters as a whole node set, the process

of feature selection can be implemented as follows: (a) clustering all the process parameters and (b) selecting representative ones for each group. Some researchers have explored the centrality and influence indicators in complex networks to reflect the importance of nodes in the network [21]. The patterns among nodes, including the differences and connections, can also be studied to find the key network participants [22]. However, the key parameters selected based on experience virtually ignore the parameter interactions such as the similarity between them and their importance in the network. Moreover, many feature extraction methods transform the original data set to another by recombining existing features into new features, which may destroy the original physical structure of data and cause the new features to lose their physical meaning. Therefore, based on the characteristics of the steel product data set, all variables can be clustered according to the correlation coefficient, and the relationships between them can be measured by the centrality and influence indicators, so as to complete the feature selection and obtain the input parameters for the subsequent learners.

With the continuous development of data mining technology, artificial intelligence methods such as neural network [23], fuzzy control [24], and expert system [25] have become more and more popular. Among them, support vector machine (SVM) is an efficient learning machine based on statistical learning theory and structural risk minimization principle proposed by Vapnik. It can deal with problems with multiple input and single output. However, problems in the steel production process often have multiple outputs which are not mutually independent. If multiple support vector machine regression (SVR) algorithms are used to estimate multiple output functions, each sample point cannot be treated equally, so the accuracy is poor. Therefore, in order to improve the accuracy of estimation and reduce the computational workload of multidimensional regression problems, multi-output support vector machine regression (MSVR) can be used for performance prediction for the steel products [26].

Motivated by the above considerations, we propose a novel prediction model for steel mechanical properties, with MSVR based on MIC and complex network clustering. In our model, we measure the correlation between features with MIC, employ hierarchical clustering analysis based on the complex network theory, quantitatively evaluate each feature by centrality and influence indicators, then choose a feature subset as a parameter input which could represent a large amount of information. The MSVR is used to predict the mechanical properties and its accuracy can verify our proposed framework. By the case analysis of the practical steel production data in a steel company in Central China, we compare our method with the full-parameter subset input, empirical subset input, and Pearson-based subset input. It turns out that our scheme has the lowest computational complexity and the highest prediction accuracy.

The remaining sections of this article are organized as follows: preliminaries about the correlation evaluation index, theory of complex network, and the performance prediction model are briefly introduced in Section 2; in Section 3, the detailed development of the proposed novel prediction model with MSVR based on MIC and complex network clustering is presented; in Section 4, an actual case of steel production is studied and the comparison analyses of prediction results are provided; and Section 5 gives conclusions.

2. Preliminaries

2.1. Correlation Analysis Methods

Correlation analysis is a basic issue in statistics that aims to quantify the association between two variables from limited data, which can be divided into linear and nonlinear. Linear correlation refers to the case that the output and input are in positive proportion or inverse proportion. When two variables share a linear relationship, the Pearson correlation is the standard measure of dependence, while it is not applicable when relationships are highly nonlinear. The nonlinear correlation is more complex and may be formed by the superposition of a variety of complex functional relationships. Therefore, it is natural to ask how to measure statistical correlation in a way that treats relationships of different types equally.

As is well known, mutual information (MI) is already widely employed to quantify associations no matter what relationship types [27]. Even though it was proposed in the communications systems, MI has been repeatedly proved to be applicable in various statistical problems. In units known as "bits", MI strictly determines how much information one variable reveals about another. The MI between two random variables *X* and *Y* is defined in terms of their joint probability distribution p(X, Y) as

$$I(X,Y) = \iint p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} dxdy$$
(1)

On the basis of MI, Reshef et al. proposed the concept of maximal information coefficient (MIC), a statistic measure other than a dependence one [18]. Compared with MI, *MIC* captures a wider range of associations both functional and not. In principle, *MIC* is based on the idea that if there is a certain relationship between two variables, a grid can be drawn on the scatter diagram of the two variables and the data can be partitioned to encapsulate this relationship. Indeed, to calculate the *MIC* of two variables, explore all grids at the maximum resolution and calculate the largest possible mutual information. Therefore, the heart of *MIC* is a naive mutual information estimate I(x, y) computed using a data-dependent grid scheme. Let *x* and *y* respectively denote the number of bins imposed on the *x* and *y* axes. The *MIC* grid scheme is chosen so that (i) the total number of bins xy does not exceed some user-specified value *B* and (ii) the value of the ratio where $Z = log_2(min(x, y))$ is maximized.

The ratio computed using this data-dependent grid scheme is how MIC is defined.

$$MIC(X,Y) = max\left\{\frac{I(x,y)}{Z}\right\} = \frac{I(x,y)}{\log_2(min(x,y))}, x \times y < n^6$$
(2)

Note that $B = n^6$. MIC(X, Y) is always nonnegative and MIC(X, Y) = 0 only when X and Y are mutually independent. Besides, MIC values will be greater than zero when X and Y show any correlations, regardless of how nonlinear that relationship is. Moreover, the stronger the correlation is, the larger the value of MIC(X, Y).

2.2. Complex Network Theory

A network consists of nodes that represent individual entities and links between each other. Actually, whether you realize it or not, we are surrounded by all kinds of networks, including transportation networks, social networks, and manufacturing networks; building networks are a good way of modeling. Based on the findings that a scale-free network has the outstanding features of strong connectivity and survivability, Barabâsi and Albert have further developed for network science a tool called complex network theory to study the topology for networks [28]. We have noticed that increasing network sizes and nontrivial topological structures concur with the increasing richness and variety of attribute information associated with the nodes in network.

Complex network is a kind of abstract model which maps the real complex system. It abstracts the entities in the complex system into nodes and the relationships between entities into lines. It can be divided into weighted network and unweighted network. The former has a binary nature where the edges between nodes are either present or not, while the latter displays a large heterogeneity in the capacity and the intensity of the connections. The adjacency matrix is a binary square matrix with the same row and column label, which is commonly used to represent the actual relationships and construct a complex network. Complex network theory is widely used to study the characteristics of various networks and further improve the network performance. The relationship between nodes in the network can be quantitatively studied by centrality analysis, binary relationship research, block-modeling analysis, and cohesive subgroup analysis, etc. [29,30].

2.2.1. Complex Network Clustering

Clustering, also known as transitivity, is a typical property of complex networks, where two nodes associated with a common node are likely to be similar. White et al. (1976) proposed the block-modeling theory [31], which can simplify the complex network according to the degree of associations between nodes. Specifically, the nodes are rearranged into blocks by clustering, and the basic characteristics of the whole network can be reflected by each block. Recently, some scholars combined the stochastic block model with clustering to define the relationship between nodes and find subgroups [32,33].

In particular, the first step of block-modeling is to partition the actors, that is, to divide them into different groups based on methods of clustering and scaling. In particular, the Convergent Correlation (CONCOR) procedure is a method of hierarchical clustering for relational data which begins by forming a new square matrix of product–moment correlations between the columns (or rows) of the original data and is found to give results that are highly compatible with analyses and interpretations of the same data using the block-modeling approach [34]. CONCOR is an iterative convergence algorithm, which measures the network structure by repeatedly calculating the correlation matrix. Each iteration of CONCOR contains a hierarchical clustering to achieve partition. According to the correlation matrix between nodes, the data set is divided into different levels and can obtain the tree clustering structure. CONCOR is an iterative convergence algorithm, which measures the network structure by repeatedly calculating the correlation matrix. Each iteration of CONCOR contains a hierarchical clustering to achieve partition. According to the correlation matrix between nodes, the data set is divided into different levels and can obtain the tree clustering structure. We repeatedly calculating the correlation matrix. Each iteration of CONCOR contains a hierarchical clustering to achieve partitions.

The purpose of complex network clustering is to find the subgroups existing in the whole network. According to the correlation, the nodes with high degree of similarity are automatically clustered into one group. Selecting the representative nodes for each group based on the importance and power indicators and eventually forming a representative node set will be better than picking up typical nodes in the whole network. The partition process of the block model is shown in Figure 1, where we can see that several scattered nodes are divided into 16 clusters according to their similarity. The similarity between nodes in one cluster is high, and the importance of each node can be evaluated.



Figure 1. Partition example of block model.

2.2.2. Centrality Evaluation of Nodes in the Complex Network

In the complex network, how to judge the power and importance of each node mainly depends on its centrality and influence. Based on the actual relationship data, we measure

the "power and status" of nodes by the following four commonly used indicators, namely degree, closeness, betweenness, and katz.

Degree Centrality

Degree centrality is defined as the number of links incident upon a node. If the network is directed, then two separate measures of degree centrality are defined, namely, in-degree and out-degree. In-degree is a count of the number of ties directed to the node and out-degree is the number of ties that the node directs to others. In many cases, the degree is the sum of in-degree and out-degree. This index reflects the "power" of a node in the network and nodes with high degree are more likely to be the center of the network.

Betweenness Centrality

Betweenness centrality is a way of detecting the amount of influence a node has over the flow of information in a graph. It is often used to find nodes that serve as a bridge from one part of a graph to another. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex. This index measures the ability of controlling the resources of each actor. If an actor is on the shortest path of many other actor-pairs, its degree is generally low, but it may play an intermediary role so as to be the center of the network.

Closeness Centrality

Closeness centrality is a way of detecting nodes that are able to spread information very efficiently through a graph. The closeness centrality of a node measures its average farness to all other nodes. Nodes with a high closeness score have the shortest distances to all other nodes. This index reflects the inverse distance of nodes to other points. If an actor is closer to other actors, it is easier to transmit information; therefore, it is more likely to be the center of the network.

Katz Centrality

In graph theory, the Katz centrality is used to measure the relative degree of influence of an actor within a social network. Unlike typical centrality measures which consider only the shortest path between a pair of actors, Katz centrality measures influence by taking into account the total number of walks between a pair of actors. Katz centrality computes the relative influence of a node within a network by measuring the number of the immediate neighbors and also all other nodes in the network that connect to the node under consideration through these immediate neighbors. This index considers the direct and indirect relationship between node and other nodes. The shorter the distance between node *i* and node *j*, the greater the impact of node *i* on node *j*.

2.3. Mechanical Property Prediction Model

2.3.1. Support Vector Regression

In contrast to simple linear regression, SVR gives us the flexibility to define how much error is acceptable in our model and will find an appropriate line (or hyperplane in higher dimensions) to fit the data. Specifically, set a threshold α and calculate the loss of data points when $|f(x) - y| > \alpha$, supposing that the data points within the threshold are predicted accurately. One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. Additionally, it has excellent generalization capability, with high prediction accuracy.

The objective function of SVR is to minimize the coefficients—more specifically, the error term, which is instead handled in the constraints, where we set the absolute error less

than or equal to a specified margin, called the maximum error, ϵ (epsilon). We can tune epsilon to gain the desired accuracy of our model.

Suppose that $x \in \mathbb{R}^d$, $y_i \in \mathbb{R}$, y_i is the output of x_i , d is the dimension, and l is the number of samples. Given the training set $\{(x_i, y_i)\}_{i=1}^l$, the goal of SVR is to find an optimal equation f from the set of hypothesis equations by minimizing the error term. The optimal equation f is as follows, where w is the weight vector and b is the threshold.

$$\left\{ f \middle| f(x) = w^T x + b, w \in \mathbb{R}^d, b \in \mathbb{R} \right\}$$
(3)

2.3.2. Multidimensional Support Vector Regression

Assume $Y = \{y_1, y_2, y_3...\}$ is the quality index set of steel products and $X = \{X^B, X^C, X^R\}$ is the process parameter set from three stages: smelting, continuous casting, and rolling. Each stage consists of many specific process parameters, such as $X^B = \{x_1^B, x_2^B, ..., x_a^B\}$, which means the number of variables in the steelmaking stage is *a*. The mean absolute percentage error (MAPE) can be set as the algorithm evaluation index, and the quality modeling considering the effect of process parameters on quality index can be abstracted as:

$$Y^{T} = \{y_{1}, y_{2}, y_{3} \dots\}^{T} \approx f(X^{B}, X^{C}, X^{R})$$
(4)

3. Mechanical Prediction Model with MSVR Based on MIC and Complex Network Clustering

3.1. Problem Description

The data collected from the intricate process of steel production are high-dimensional and coupled with each other. There are complex linear or nonlinear relationships between them; meanwhile, their impact on product quality is hereditary. If we use the full-parameter data to model, not only is the calculation complex, but also the modeling is often inefficient and cannot well reflect the real problem because of the redundant features. If the important and representative features can be selected from the high-dimensional process data to simplify the complex problem, the subsequent modeling will be simpler and the effect will be more obvious. The emphasis of this paper is how to select the representative feature subset from the full-feature set, and then predict the performance of steel products more accurately.

Taking the throughout process of steel production as an example, after cleaning and deduplication of the original data set O, which means removing the parameters that are completely irrelevant to the mechanical properties, namely the MIC between them being less than 0.05. The left process parameters from three typical stages, namely, steelmaking, continuous casting, and rolling are defined as $F = \{X^B, X^C, X^R\} = \{x_1, x_2, x_3 \dots, x_m\}$ and the number set of each stage is {a, b, c}, which means that the number of total parameters is (a + b + c = m) [35]. Define the mechanical property set $Y = \{y_1, y_2, y_3\}$ which contains three indicators: tensile strength, yield strength, and elongation. The purpose of this study is to use a certain feature selection method to obtain representative and low-dimensional feature subset $X = \{x_1, x_2, x_3 \dots, x_t\}, t \ll a + b + c$ from high-dimensional variable set $\{X^B, X^C, X^R\}$, and perform the subsequent MSVR performance prediction modeling $Y^T = f(x_1, x_2, x_3 \dots, x_t)$, which could effectively simplify the calculation and improve prediction accuracy at the same time.

3.2. Model and Algorithm

Based on the relevant basic theories in Section 2 and the requirements in Section 3.1, we propose an algorithm that firstly uses MIC to measure the linear and nonlinear correlation relationships between high-dimensional parameters. Secondly, we construct a complex network, and quantitatively evaluate each feature by CONCOR clustering method and centrality and influence analysis. Eventually, we could obtain the feature subset that could represent the full parameters efficiently, which could be used as the input of MSVR and to predict the mechanical properties. In order to verify the effectiveness and feasibility of the

algorithm, the full-parameter set, empirical subset, and the best feature subsets selected based on MIC and Pearson coefficients are used as input for MSVR respectively, and the method with the least error and the optimal feature subset could be obtained.

The model and algorithm of this paper can be divided into two parts. One is the prediction model based on MSVR, the other is the feature selection algorithm based on correlation measurement and complex network, as shown in Figure 2.



Figure 2. The prediction model for steel mechanical properties with multidimensional support vector regression (MSVR) based on maximum information coefficient (MIC) and complex network clustering.

3.2.1. Correlation Measurement

Suppose that $C(x_i, x_j)$ is the correlation coefficient between x_i and x_j . In this paper, MIC is used to measure the linear and nonlinear correlation between attributes. In order to verify the representation effect of MIC, the Pearson coefficient between attributes is also calculated for modeling.

Create the correlation matrix *C* by the correlation coefficient between features, and construct the complex network that characterizes the correlation between features. This matrix is a symmetric matrix with diagonal 1.

$$C = \begin{bmatrix} 1 & C(x_1, x_2) & C(x_1, x_3) & \dots & C(x_1, x_m) \\ C(x_2, x_1) & 1 & \cdots & \dots & C(x_2, x_m) \\ C(x_3, x_1) & \vdots & \ddots & \dots & \vdots \\ \dots & \dots & \dots & 1 & \dots \\ C(x_m, x_1) & C(x_m, x_2) & \dots & \dots & 1 \end{bmatrix}$$
(5)

3.2.2. The Clustering Model Based on the Complex Network and CONCOR Algorithm

A complex network is constructed based on the correlation matrix *C* and the CONCOR algorithm is employed to build the block model. The CONCOR algorithm calculates the Pearson correlation coefficient of the correlation matrix iteratively and carries out the hierarchical clustering, starting from the initial correlation matrix. The flow of the algorithm is shown in Algorithm 1. After the CONCOR, the partition of features is realized. Define the subgroup as $G = \{g_1, g_2, ..., g_t\}$ where *t* is the number of subgroups, and $g_i = \{x_1, x_2, ..., x_j\}, i \leq t, j \ll a + b + c$, where *j* is the number of features in the subgroup g_i .

Algorithm 1. The clustering model based on CONCOR.

Input : correlation matric C1 and the partition level at which any pair of actors is aggregated. Output : C2, which denotes the correlation coefficient matrix of C1 and blocks represented in terms of a clustering dendrogram clustering graph under different levels.

Step 1 : Calculate C2 which is the Pearson correlation coefficient of C1.

Step 2 : The blocks are given for each level at which any pair of actors is aggregated. Carry out the hierarchical clustering from the max level, and combine two features with the highest similarity. The similarities of partitions from the same level should all reach one corresponding value, and one feature can only exist in one group.

Step 3: Reduce level by 1, which means reducing the corresponding similarity value of clusters, and look for the features with highest similarity to the clustered partitions from the unclustered features, which could cluster by themselves, or be added into the existing partition. Step 4: Iterate Step 3 until level = 1 when all features enter the same group.

3.2.3. Feature Evaluation

Given the different partitions of different levels and the similarity complex network, we comprehensively evaluate the nodes in each subgroup for feature selection with four centrality and influence indicators that we mentioned before. What should be pointed out is that the measure of degree centrality is based on the weighted matrix which is the initial correlation matrix, while the measures of betweenness, closeness, and Katz centrality are based on the unweighted matrix which is the binarization of initial correlation matrix. Suppose that *n* denotes the number of nodes in the network.

Degree Centrality

The absolute degree $C_{AD}(x)$ is the sum of the weights between node x and all other nodes, and the relative degree $C_{RD}(x_i)$ is the absolute centrality divided by the maximum possible degree (n - 1).

$$C_{RD}(x_i) = \frac{\sum_j C(x_i, x_j)}{n-1} \tag{6}$$

Betweenness Centrality

Define g_{jk} as the geodesic between node j and k, and $g_{jk}(x_i)$ as the number of geodesics that go through the node x_i . The absolute betweenness centrality $C_{AB}(x_i)$ is the sum of the probabilities that node x_i is on the shortest path between all pairs of points. The rela-

tive betweenness $C_{RB}(x_i)$ is the absolute betweenness divided by the maximum possible betweenness $(n^2 - 3n + 2)/2$.

$$C_{AB}(x_i) = 2\sum \frac{g_{jk}(x_i)}{g_{jk}} / \left(n^2 - 3n + 2\right)$$
(7)

Closeness Centrality

Define $Farnessx_i$ as the sum of the geodesic distances between node x_i and all other nodes, d_{ij} as the geodesic distances between node x_i and x_j , and the absolute closeness centrality C_{APi} is the reciprocal of $Farnessx_i$. The relative closeness centrality $C_{RP}(x_i)$ is $C_{AP}(x_i)$ divided by the maximum possible closeness 1/(n-1).

$$C_{RP}(x_i) = \frac{\frac{1}{Farness(x_i)}}{\frac{1}{n-1}} = \frac{n-1}{Farness(x_i)} = \frac{n-1}{\sum_{j=1}^n d_{ij}}$$
(8)

Katz Centrality

Katz centrality measures the influence by considering the direct and indirect support or attention between nodes. Define *S* as a matrix consisting of 0 and 1 that reflects the direct-connection relationships between actors when the path length is 1, and $S_{ij} = 1$ denotes that the actor *j* connects to actor *i* directly and the length is 1. The sum of *j*-column represents the total number of times that actor *j* connects to other actors by 1; define S_{ij}^2 as the number of paths that connect the actor *i* and *j* by length 2 and S_{ij}^3 by length 3, and so on. Considering that the higher the power of the matrix S_{ij}^* , the lower the effect of the influence, so an attenuation factor α is introduced to characterize this performance. The value of α depends on the situation and $1/a \in (b, 2b)$. When $\alpha = 0$, it decays completely and when $\alpha = 1$, it does not. For a matrix where the elements are nonnegative, *a* simple upper limit of the maximum eigenvalue *b* is the maximum sum of rows.

Define P = [Degree, Betweenness, Closeness, Katz]. In order to eliminate the influence of dimension, we sort the four indicator values and get four ranking values to measure their comprehensive centrality and influence. Define $R = [R_D, R_B, R_C, R_K, R_T]$ where R_D, R_B, R_C, R_K, R_T represent the ranking values of four centrality indicators and the total ranking respectively.

$$R_T = R_D + R_B + R_C + R_K \tag{9}$$

3.2.4. Feature Selection

Suppose that $R_T^i = \{R_T^{i_1}, R_T^{i_2} \dots R_T^{i_p}\}$ is the total ranking matrix of the features in the subgroup g_i , where p denotes the feature number of g_i . Select the feature with the top total ranking as the subgroup representation, namely $R_T^{i_1} = minR_T^i$. In this way, explore all subgroups and obtain the feature set $\{x_1, x_2, x_3 \dots, x_t\}$, where t is the number of subgroups.

3.2.5. Mechanical Property Prediction Based on MSVR

The above work can obtain the feature selection results respectively based on the MIC and Pearson correlation characterization. Moreover, in order to verify the effect of our proposed method, the empirical subset and full-parameter subset are used for comparative experiments. Applying the above four feature sets to construct the data set for MSVR modeling, we divide the training set and the test set, and perform cross validation test to verify the error. It should be pointed out that even though we are using the same correlation characterization, different partition levels get different feature selection results, corresponding to different MSVR prediction results.

4. Case Study and Discussion

In order to test the feasibility and efficiency of the proposed prediction model, we collected a total of 1607 data samples of the whole production process from a steel company in Central China and verified our model. The product is the cold-rolled strip and the steel

grades selected in our experiment include DR01, DR02, DR04, DR06, DX51, DX52, DX53, SPCC, SPCD, SPCE, SPCF, SPCG, etc. The data come from four main processes: smelting, continuous casting, hot rolling, and cold rolling. The original parameters influence each other and contain a lot of linear and nonlinear relationships, of which the number is 211. The deduplication process is described as follows: calculate the MIC values between the original parameters and three mechanical properties, and remove the ones completely irrelevant to properties, namely the MIC between them is less than 0.05. Finally, a total of 111 process parameters were obtained as the full-parameter subset. The number of parameters in each process stage is shown in Table 1. The exactly defined chemical compositions include C, Si, Mn, P, S, Ni, CR, Cu, ALS, ALT, AS, B, MO, N, NB, PB, SN, TI, which are all collected in the smelting stage. Among the 1607 data samples, the average, maximum value, and variance of the chemical composition contents are listed in Table 2.

Table 1. Number of parameters in each process stage.

Process Stage	Number of Process Parameters
Smelting	18
Continuous casting	32
Hot rolling	42
Cold rolling	19
Sum	111

Table 2. The average and variance of the chemical composition contents.

Composition	Maximum	Average	Variance	Composition	Maximum	Average	Variance
С	0.203	0.011914	0.020479	ALT	0.245	0.052757	0.028082
Si	0.1299	0.008354	0.006990	AS	0.0245	0.005758	0.002594
Mn	0.8415	0.170039	0.081668	В	0.0009	0.000213	0.000183
Р	0.071	0.015293	0.006540	MO	0.0089	0.001468	0.000834
S	0.0335	0.008336	0.004066	Ν	0.0074	0.001969	0.001137
Ni	0.038	0.00974	0.004460	NB	0.0374	0.000866	0.001014
CR	0.0684	0.020961	0.010005	PB	0.0025	0.000771	0.000296
Cu	0.084	0.019762	0.010511	SN	0.0192	0.00617	0.002472
ALS	0.2235	0.050082	0.026820	TI	0.3064	0.042642	0.042190

4.1. Correlation Calculation and Partition Results

The distribution of MIC values among the 111 process parameters is shown in Figure 3. It can be seen that nearly 50% of MIC values are greater than 0.43 and 34 values are more than 0.8, which indicates that there are indispensable correlation relationships between these features. It is necessary to mine these relationships and remove redundant features, so as to clarify the nature of the relationships between features and simplify the input data set of subsequent modeling.

We construct a complex network based on the MIC matrix, and carry out the CONCOR to build a block model. Set the initial clustering level as 4, and Figure 4 shows the number of partitions under different clustering levels. It can be seen that the number of partitions gradually increases with the rise of clustering level, and the clustering stops when the clustering level is 9, meanwhile the number of partitions is the maximum, 71.

Combined with the partition results shown in Figure 5 of which the clustering level is 4 to 9 respectively, it can be seen that the higher the level, the more partitions. This is because the next level of clustering is based on the previous level, which means expanding the feature numbers within a partition by reducing the similarity of the group, so the number of partitions will decrease. The first clustering level is 9, then the next clustering is based on level 9 which expands the members of each group and reduces the partition number. When the clustering level is 1, all features are in the same partition.



Figure 3. MIC value distribution among full process parameters.



Figure 4. The corresponding relation between the clustering level and the number of partitions.

A partition at level 6 is used to show the clustering process. Define $\{a : b\}$ or $\{a\}$ as the process parameter information where *a* denotes the serial number and *b* denotes the name. Starting from level 9, {72: RF_IN_TT, 73: RF_EX_TT} and {8: COIL_THK_MAX, 9: COIL_THK_MIN} are assigned to a respective partition firstly because the MIC values between them are 0.9685 and 0.9759 separately, which are the highest. When the level is 7, {1: THK_ACT} joins the partition {8: COIL_THK_MAX, 9: COIL_THK_MIN} because the MIC values between them are both 0.998706, which is the highest. In the same way, when the level is 6, {1: THK_ACT, 8: COIL_THK_MAX, 9: COIL_THK_MIN} and {72: RF_IN_TT, 73: RF_EX_TT} are clustered into a larger partition. The dendrogram is shown in Figure 6, and the MIC values between 5 parameters are shown in Table 3.



Figure 5. Partition results under different clustering levels.



Figure 6. Clustering process of feature 1, 8, 9, 72, 73 (the number in the ellipse is the maximum information coefficient (MIC) value between related parameters).

Table 3. Maximum information coefficient	(MIC)	values between	features 1	, 8,	9,7	2,73
--	-------	----------------	------------	------	-----	------

Feature Name	Serial Number	1	8	9	72	73
THK_ACT	1	1	0.9987	0.9987	0.8318	0.7765
COIL_THK_MAX	8	0.9987	1	0.9759	0.8926	0.8547
COIL_THK_MIN	9	0.9987	0.9759	1	0.8714	0.7941
RF_IN_TT	72	0.8318	0.8926	0.8714	1	0.9685
RF_EX_TT	73	0.7765	0.8547	0.7941	0.9685	1

4.2. Feature Evaluation and Selection

As mentioned above, four centrality and influence indicators are selected to evaluate the importance of each parameter and we rank them by category. Table 4 shows the top 20 features with the highest total ranking and their respective rankings of the four indicators. Table 5 shows the detailed information of the top 20 features including the feature name and the cluster number using the MIC-based model at level 4. The last column indicates whether the feature is selected in its cluster. It can be found that the five rankings are highly related. The parameters with the high total rankings tend to rank at the top of the four separate indicators. Among them, the process parameter "TI" ranks respectively 1, 9, 1, 1

at degree, betweenness, closeness, and Katz centrality and the total ranking is 12, which means that this feature owns greater power and is the most representative in the partition.

Feature	nDegree	D-Rank	nBetweenness	B-Rank	nCloseness	C-Rank	Katz	K-Rank	Total Ranking
36	0.51	1	0.02	9	0.14	1	3.32	1	12
81	0.50	2	0.03	8	0.14	7	3.14	4	21
82	0.48	5	0.03	3	0.14	10	3.07	9	27
64	0.45	11	0.05	1	0.14	2	3.01	14	28
6	0.47	7	0.01	16	0.14	8	3.12	6	37
89	0.47	6	0.01	17	0.14	11	3.12	5	39
70	0.43	24	0.02	11	0.14	3	3.20	3	41
105	0.43	23	0.02	12	0.14	4	3.20	2	41
65	0.44	14	0.03	4	0.14	5	2.95	20	43
22	0.46	8	0.02	10	0.14	9	2.95	19	46
66	0.43	21	0.03	5	0.14	6	2.86	23	55
74	0.48	4	0.01	29	0.14	15	3.07	8	56
94	0.43	22	0.01	18	0.14	13	3.08	7	60
79	0.48	3	0.00	33	0.14	17	3.04	10	63
96	0.43	26	0.01	23	0.14	18	3.02	13	80
83	0.46	10	0.01	30	0.14	23	2.95	18	81
2	0.40	43	0.01	24	0.14	14	3.02	11	92
28	0.40	39	0.03	6	0.14	19	2.71	32	96
47	0.38	52	0.04	2	0.14	12	2.71	31	97
90	0.40	44	0.01	25	0.14	16	3.02	12	97

Table 4. The top 20 features with the highest total ranking and their respective rankings of the four indicators.

Table 5. The detailed information of the top 20 features with the highest total ranking.

Feature Number	Feature Name	Stage	Cluster Number	Selected
36	TI	Smelting	3	1
81	CT_TT	Hot rolling	4	1
82	FET_PT	Hot rolling	4	0
64	IDN_MIN	Continuous casting	11	1
6	COIL_WID	Cold rolling	6	1
89	CWC_1	Hot rolling	6	0
70	SLAB_LEN	Hot rolling	6	0
105	IM_LEN_1	Hot rolling	6	0
65	IDN_MAX	Continuous casting	11	0
22	MN	Smelting	11	0
66	IND_AVE	Continuous casting	11	0
74	FET_TT	Hot rolling	4	0
94	WID_AVE	Hot rolling	6	0
79	FDT_TT	Hot rolling	4	0
96	WID_MIN	Hot rolling	6	0
83	FET_NT	Hot rolling	4	0
2	MAT_WID_ACT	Cold rolling	5	1
28	ALS	Smelting	12	1
47	ST_AVE	Continuous casting	13	1
90	CWC	Hot rolling	5	0

Finally, the feature selection is based on the partition situation and the feature evaluation results. At different clustering levels, compare the centrality and influence rankings of different features in each partition. The top 1 feature is selected as the representative of the partition, also as a member of the selected feature subset. For example, when the clustering level is 4, 111 features are divided into 16 subgroups. The feature distribution of the first subgroup g_1 is shown in Figure 7.

$$R_T^1 = \left\{ R_T^{1\,1}, R_T^{1\,2} \dots R_T^{1\,15} \right\} = \left\{ 240, 160, 226, 247, 175, 191, 183, 187, 191, 187, 167, 159, 165, 194, 196 \right\}$$
(10)



Figure 7. The individual network of subgroup g_1 when the clustering level is 4.

There are 15 features in subgroup g_1 , and Figure 8 shows the total ranking scatter diagram of each feature. It can be seen that the ranking distribution within the subgroup is relatively concentrated among [150, 200], which also verifies the rationality of clustering, that is, the rankings of similar features should also be similar. Among them, the top ranking is feature 93 whose total ranking is 159, so select feature 93 as the representative feature of subgroup g_1 and add it into the final feature subset.



Figure 8. The total ranking of each feature in subgroup g_1 when the clustering level is 4.

The rest can be deduced by analogy. Select the representative features of all partitions at level 4, and then expand the clustering level. Finally, the feature subsets at level 4–9 are obtained, as shown in Table 6.

Compared with level 8, level 9 has two new representative features, which are {24: S} and {38: PS_MIN}. This is because at level 8, {24} enters the partition {23,25} with the corresponding correlation coefficient {0.5175,0.7232}, 39 enters {38,40} with {0.9021,0.9749}. Both of them are the closest feature to the corresponding partition, and the process conforms to the clustering rules mentioned before.

In addition, we can discover that the clustering at level 9 is most concise with the least feature numbers and the highest correlation in each partition. Therefore, it can be estimated that the representative features selected at this level may have the best prediction effect, which can also be proved in the follow-up article.

Level	Feature Number	Added Feature Number	Added Feature Subset
4	16		{93,7,36,81,2,6,10,86,17,67,64,28,47,43,56,59}
5	31	15	{104,80,101,79,107,70,12,34,69,18,30,53,48,54,62}
6	49	18	{72,75,14,4,103,84,71,111,94,22,26,31,39,51, 42,48,50,57,61}
7	62	13	{8,5,16,3,98,82,95,66,35,21,32,41,46}
8	69	7	{1,92,99,74,78,23,63}
9	71	2	{24,38}

Table 6. Feature subset selection at cluster level 4–9.

4.3. MSVR Property Prediction Model

According to the feature selection results, the original sample data are divided into the training set and test set at the ratio of 8:2 to train the MSVR model. Three mechanical properties are selected, which are lower yield strength, tensile strength, and elongation, separately. The mean absolute percentage error (MAPE) is chosen as the evaluation index of the effectiveness of the proposed algorithm. We calculate three MAPE values and the average of them to represent the prediction accuracy. We choose four parameter sets as the input, which are MIC-based subset, Pearson-based subset, full-parameter subset, and empirical subset to perform the comparison experiment.

MSVR Prediction Results at Different Clustering Levels

Figure 9 shows the MAPE comparison between feature selection results based on MIC-based subset, full-parameter subset, and the empirical subset. It can be seen that starting from level 5, four kinds of prediction error (including three MAPE values of three mechanical properties and their average) of our proposed algorithm are all lower than the other two input sets. In addition, as the number of selected features increases from level 4 to level 9, the growth rate slows down and the prediction error decreases. At level 9, the number of features reaches the maximum 71, while the four MAPE values all reach the lowest.



Figure 9. Error comparison between MIC-based feature selection and full-parameter, empirical subset modeling.

As shown in Figure 10, when the clustering level is 9, the prediction error of the optimal feature subset is significantly lower than that of the full-parameter and empirical subset. Therefore, it can be concluded that the feature selection method proposed in this paper can select a small number of parameters from the full-parameter set to represent the whole, and the prediction effect is better.



Figure 10. Prediction error comparison between the optimal subset of MIC and full-parameter, empirical subset.

In order to verify that the MIC-based feature selection method can characterize the nonlinear correlation relationships between features more reasonably, we use the Pearson coefficient matrix to represent the initial correlation, and compare the prediction error under two correlation measures. The prediction error of "lower yield strength" (left) and the average error of three mechanical properties (right) are shown in Figure 11. It can be seen that the overall error of MIC-based feature selection method is lower than that of Pearson-based method. With the increase of clustering level, the prediction accuracy difference between the two methods gradually becomes smaller.



Figure 11. Comparison of prediction errors of feature selection methods based on MIC and Pearson coefficient. (**a**) The prediction error of "lower yield strength". (**b**) The average error of three mechanical properties.

When the clustering level is 4, the prediction accuracy of MIC method is 1.69% higher than that of Pearson and only one feature 10 is coincident, which means that the two similarity measurement methods are quite different, among which MIC is better. Apparently, compared with the Pearson coefficient, MIC can widely explore the linear and nonlinear relationship between process parameters.

To sum up, the feature selection method based on MIC and complex network clustering can represent the global situation with fewer features and better prediction effect than the full-parameter subset. At the same time, compared with the empirical subset and the similarity measurement based on Pearson coefficient, our model has higher prediction accuracy. It should be pointed out that no matter whether the feature selection is based on MIC or Pearson coefficient, the prediction accuracy is both higher than that of fullparameter subset, which indicates that there are a lot of linear and nonlinear relationships in the original data set. If it can be well mined and analyzed, the difficulty of subsequent modeling can be greatly reduced.

5. Conclusions

Aiming at the complex industrial process of steel production, this paper proposes a property prediction model based on MIC and complex network clustering, which adopts the MSVR on the basis of attribute selection. Compared with full-parameter subset, empirical subset, and feature selection subset based on Pearson coefficient, our scheme has the lowest computational complexity and the highest prediction accuracy.

The innovation and research significance of this paper are as follows:

- 1. The feature selection method based on MIC and complex network theory can effectively solve the attribute reduction problem of complex process data. The MIC-based subset input has higher prediction accuracy compared with the Pearson-based subset, full-parameter subset, and empirical subset input. Specifically, the average prediction errors of the three mechanical properties with the four different inputs are 2.359%, 2.428%, 2.872%, and 3.010%, respectively;
- 2. Using MIC to measure the similarities between process parameters can mine many linear and nonlinear relationships, including all kinds of interesting correlations, which are strongly meaningful in constructing the relationship complex network;
- 3. The centrality and influence theory of complex network can be used to measure the importance of process parameters in the network efficiently;
- 4. The feature selection method proposed in this paper reduces the calculation complexity and simplifies the actual problem. The selected feature subset still has physical meaning and does not destroy the original feature structure. At the same time, the average error of the MSVR prediction model is also greatly reduced.

Author Contributions: Conceptualization, Y.W. and Z.L.; data curation, Y.W. and Z.L.; funding acquisition, Z.L.; methodology, Y.W. and Z.L.; software, Y.W. and Y.Y.; validation, Y.W. and Y.Y.; visualization, Y.W.; writing—original draft, Y.W.; writing—review and editing, Y.W. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Fundamental Research Funds for the Central Universities, grant number FRF-MP-20-08.

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

References

- 1. Qiao, L.; Wang, Z.; Zhu, J. Application of improved GRNN model to predict interlamellar spacing and mechanical properties of hypereutectoid steel. *Mater. Sci. Eng. A* 2020, 792, 139845. [CrossRef]
- Sui, X.Y.; Lv, Z.M. Prediction of the mechanical properties of hot rolling products by using attribute reduction ELM. *Int. J. Adv. Manuf. Technol.* 2016, 85, 1395–1403. [CrossRef]
- 3. Gao, H.; Xu, Y.; Zhu, Q. Spatial Interpretive Structural Model Identification and AHP-Based Multimodule Fusion for Alarm Root-Cause Diagnosis in Chemical Processes. *Ind. Eng. Chem. Res.* **2016**, *55*, 3641–3658. [CrossRef]
- 4. Ayesha, S.; Hanif, M.K.; Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf. Fusion* **2020**, *59*, 44–58. [CrossRef]

- Cano, A.; Ventura, S.; Cios, K.J. Multi-objective genetic programming for feature extraction and data visualization. *Soft Comput.* 2017, 21, 2069–2089. [CrossRef]
- 6. Bommert, A.; Sun, X.; Bischl, B.; Rahnenführer, J.; Lang, M. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput. Stat. Data Anal.* 2020, 143, 106839. [CrossRef]
- 7. González, J.; Ortega, J.; Damas, M.; Martín-Smith, P.; Gan, J.Q. A new multi-objective wrapper method for feature selection— Accuracy and stability analysis for BCI. *Neurocomputing* **2019**, *333*, 407–418. [CrossRef]
- 8. Karasu, S.; Altan, A.; Bekiros, S.; Ahmad, W. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy* **2020**, *212*, 118750. [CrossRef]
- 9. Wu, J.-S.; Song, M.-X.; Min, W.; Lai, J.-H.; Zheng, W.-S. Joint Adaptive Manifold and Embedding Learning for Unsupervised Feature Selection. *Pattern Recognit.* 2020, *112*, 107742. [CrossRef]
- Najafi, A.; Joudaki, A.; Fatemizadeh, E. Nonlinear Dimensionality Reduction via Path-Based Isometric Mapping. *IEEE Trans.* Pattern Anal. Mach. Intell. 2016, 38, 1452–1464. [CrossRef] [PubMed]
- 11. Wilson, S.R.; Close, M.E.; Abraham, P. Applying linear discriminant analysis to predict groundwater redox conditions conducive to denitrification. *J. Hydrol.* **2018**, *556*, 611–624. [CrossRef]
- 12. Guillemot, V.; Beaton, D.; Gloaguen, A.; Lofstedt, T.; Levine, B.; Raymond, N.; Tenenhaus, A.; Abdi, H. A constrained singular value decomposition method that integrates sparsity and orthogonality. *PLoS ONE* **2019**, *14*, e0211463. [CrossRef]
- 13. Kumar, N.; Singh, S.; Kumar, A. Random permutation principal component analysis for cancelable biometric recognition. *Appl. Intell.* **2018**, *48*, 2824–2836. [CrossRef]
- 14. Narayana, P.L.; Lee, S.W.; Park, C.H.; Yeom, J.-T.; Hong, J.-K.; Maurya, A.K.; Reddy, N.S. Modeling high-temperature mechanical properties of austenitic stainless steels by neural networks. *Comput. Mater. Sci.* **2020**, *179*, 109617. [CrossRef]
- Narayana, P.L.; Kim, J.H.; Maurya, A.K.; Park, C.H.; Hong, J.-K.; Yeom, J.-T.; Reddy, N.S. Modeling Mechanical Properties of 25Cr-20Ni-0.4C Steels over a Wide Range of Temperatures by Neural Networks. *Metals* 2020, 10, 256. [CrossRef]
- Sun, L.; Yin, T.; Ding, W.; Qian, Y.; Xu, J. Multilabel feature selection using ML-ReliefF and neighborhood mutual information for multilabel neighborhood decision systems. *Inf. Sci.* 2020, 537, 401–424. [CrossRef]
- 17. Zhang, Y.; Zhu, R.; Chen, Z.; Gao, J.; Xia, D. Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data. *Eur. J. Oper. Res.* **2021**, *290*, 235–247. [CrossRef]
- Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* 2011, *334*, 1518–1524. [CrossRef] [PubMed]
- Wang, S.; Zhao, Y.; Shu, Y.; Yuan, H.; Geng, J.; Wang, S. Fast search local extremum for maximal information coefficient (MIC). J. Comput. Appl. Math. 2018, 327, 372–387. [CrossRef]
- Manbari, Z.; AkhlaghianTab, F.; Salavati, C. Hybrid fast unsupervised feature selection for high-dimensional data. *Expert Syst. Appl.* 2019, 124, 97–118. [CrossRef]
- 21. Chen, W.; Teng, S.-H.; Zhang, H. A graph-theoretical basis of stochastic-cascading network influence: Characterizations of influence-based centrality. *Theor. Comput. Sci.* 2020, 824, 92–111. [CrossRef]
- 22. Long, J.C.; Cunningham, F.C.; Carswell, P.; Braithwaite, J. Patterns of collaboration in complex networks: The example of a translational research network. *BMC Health Serv. Res.* **2014**, *14*, 1–10. [CrossRef]
- 23. Ahmadi, H.; Rezaei Ashtiani, H.R.; Heidari, M. A comparative study of phenomenological, physically-based and artificial neural network models to predict the Hot flow behavior of API 5CT-L80 steel. *Mater. Today Commun.* 2020, 25, 101528. [CrossRef]
- 24. Mohammadian, M. Modelling, Control and Prediction using Hierarchical Fuzzy Logic Systems: Design and Development. *Int. J. Fuzzy Syst. Appl.* **2017**, *6*, 105–123. [CrossRef]
- Yousefi, S.; Zohoor, M. Effect of cutting parameters on the dimensional accuracy and surface finish in the hard turning of MDN250 steel with cubic boron nitride tool, for developing a knowledged base expert system. *Int. J. Mech. Mater. Eng.* 2019, 14, 1. [CrossRef]
- 26. Shahri, A.A.; Moud, F.M.; Mirfallah Lialestani, S.P. A hybrid computing model to predict rock strength index properties using support vector regression. *Eng. Comput.* **2020**, 1–16. [CrossRef]
- 27. Kinney, J.B.; Atwal, G.S. Equitability, mutual information, and the maximal information coefficient. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3354–3359. [CrossRef]
- 28. Albert, R.; Jeong, H.; Barabási, A.-L. Error and attack tolerance of complex networks. Nature 2000, 406, 378–382. [CrossRef]
- 29. Kitsak, M.; Gallos, L.K.; Havlin, S.; Liljeros, F.; Muchnik, L.; Stanley, H.E.; Makse, H.A. Identification of influential spreaders in complex networks. *Nat. Phys.* 2010, *6*, 888–893. [CrossRef]
- 30. Ahn, Y.-Y.; Bagrow, J.P.; Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **2010**, *466*, 761–764. [CrossRef]
- 31. White, H.C.; Boorman, S.A.; Breiger, R.L. Social Structure from Multiple Networks. Amer. J. Sociology 1976, 81, 730–780. [CrossRef]
- 32. Decelle, A.; Krzakala, F.; Moore, C.; Zdeborova, L. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* **2011**, *84*. [CrossRef]
- 33. Lei, J.; Rinaldo, A. Consistency of spectral clustering in stochastic block models. Ann. Stat. 2015, 43, 215–237. [CrossRef]

- 34. Breiger, R.L.; Boorman, S.A.; Arabie, P. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *J. Math. Psychol.* **1975**, *12*, 328–383. [CrossRef]
- 35. Yan, Y.F.; An, L.D.; Lv, Z.M. Prediction of mechanical properties of cold rolled products based on maximum information coefficient attribute selection. J. Cent. South Univ. (Sci. Technol.) 2020, 051, 68–75.