Renewable Energy 63 (2014) 779–784

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Network study of the distribution of talent in the Chinese renewable energy industry

Guodong Li [a], Lin Liu [b], Han Luo [a,*]

[a] School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
[b] School of Economics and Management, North China Electric Power University, Beijing 102206, China

ARTICLE INFO

Article history:
Received 14 June 2013
Accepted 22 October 2013
Available online

Keywords:
Renewable energy talents carrier
Network
Top-k
Neighbor Rank
Gravity Rank

ABSTRACT

Because of the current energy crisis, researchers and governments pay more attention to new and renewable energy. Renewable energy is a new area of development in China, and thus many talents are attracted to pursue research in these directions with the result that the renewable energy field has an urgent need to understand how the various talents are divided within the field and the statistical findings that can be determined from this. Based on this division and the statistical information it is possible to make decisions on how to plan for the renewable energy field's development in areas, for example, such as resource delivery. To begin with we can create a statistical map of the talents' social network based on the talent carriers and their research directions. Then we use *top-k* talents to represent the network and divide the network into *k* parts. Analyzing the properties of a single point's neighbors and the transition probability of these points and those around them, we propose the value of Neighbor Rank (NR). By taking the point's gravity into account we put forward the Gravity Rank (GR). We then combine these 2 models. The experiment not only gives a reasonable result but also the *k* parts are always the *k* main areas of the renewable energy field.

1. Introduction

In the 21st century, governments are paying increased attention to the fields of new and renewable energy. In China, some state-owned companies (such as China Datang Corporation, China Huaneng Group, etc.) invest a large amount of money on renewable energy development, especially in the area of wind power. The reason for this in China is obvious. For example, because of such problems as air pollution and limited coal resources, the government sets limits on the building of thermal power plants. Along with concerns regarding nuclear leakage (citizen's affairs for the nuclear power plants), the government has a definite need to improve the deployment of new and renewable energy sources. Given the above, along with the other advantages that can be realized in pursuing renewable energy resources, many researchers spare no effort to develop this field in areas such as energy production techniques, manufacturing, and human resources, etc.

Because of the rapid development of renewable energy and future planning, governments and the researchers pay close attention to the renewable energy field's talent carriers [1,2] and the methods for determining division of talent, analysis of the various areas within the field, etc. Generally speaking, considering the difference between the renewable energy field and the more traditional energy production industry, it is difficult to determine the supply and the demand levels of talent within these industries. The development of the energy industry requires the continuous input of professional talents. Only by collecting and analyzing the information about talents in this field can the government conduct scientific planning of human resources. Therefore, a good understanding of the current situation of the talents is of the highest importance. Here, the word talent is used to refer to the professional people working in the energy field.

We are thus faced with the following problems:

(1). How to divide so many professional talents (researchers, professors, employees, etc.) into meaningful groups (parts, clusters) so that the talents who are operating in the same or similar research directions can be included into the same group.
(2). How to give a valid measurement value of the talents (ranking) so that by using this value we can assign the talents efficiently or choose the top persons in an area to conquer a problem.

* Corresponding author.
E-mail addresses: lgd@ncepu.edu.cn (G. Li), liulin0627@126.com (L. Liu), lbyzmy@163.com (H. Luo).

0960-1481/$ – see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.renene.2013.10.039

The questions above are what we want to solve. To start with, we have a database of existing talents. This information can be obtained through the government or the universities. It is obvious the size of the database is huge, so we need to divide the talents into groupings based on their research directions or their hobbies so that we will be able to analyze the information of the groups instead of the whole database. One key to this is to analyze the social networks of the talents in order to determine the effect each one has on the others and to determine which areas of the new and renewable energy fields they are most active in. Fig. 1 shows an example of some talents; each point is a talent and the borders are used to describe the associations between two talents. Fig. 1 is a network of talents and the network is built using the method of a talent vector network [1].

Because of the characteristics of the new and renewable energy industry, this article proposes an algorithmic solution to divide the talents into the various groups. Each group is a set of talents who have the similar research directions. In Fig. 1, the talents are divided into 3 parts; and it can be seen that we can use talents 11, 5 and 15 to represent the network because they have the most complex links within the network. This means these 3 talents should be assigned a high rank value.

As we have said in Part 1 of this article, we first use a vector to represent a talent with the components of the talent vector being determined by the direction of research. Then we assign the definitions of the binary relations and with these we build the network (talents directed network). Finally, the PageRank [3] algorithm is used to calculate a value which can be used to measure the importance of a talent.

2. Related work

As we mentioned above, what we want to do is to extract the *top-k* talent information from the existing social network. Commonly, the problem is one of data mining (Social Networks). Generally speaking, we have a matrix (n^*n), and after the computations we can find the value for *top-k*.

Xiaojin Zhu et al. [4] ranks the objects based on the seemingly random path within the matrix, and the rank algorithm GRASSHOPPER is based on the similarity matrix (probability's transition matrix) and then regards the matrix as an instance of the Markov Model. Finally, they rank the probabilities by determining the stationary distribution and thus determine the value for the *top-k* [4,5].

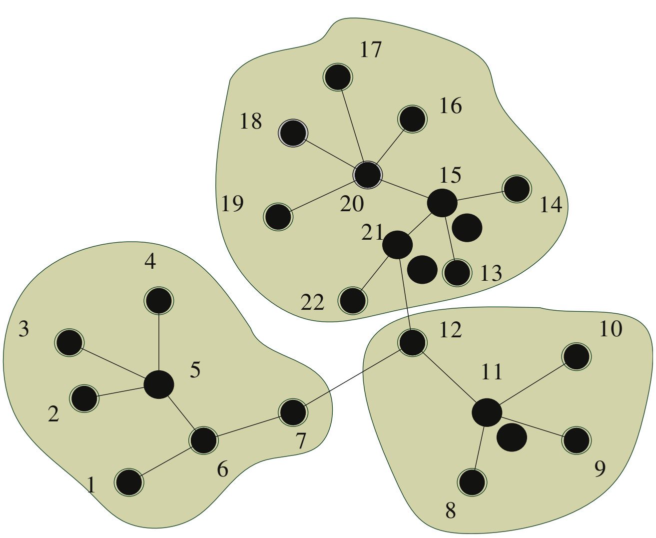

Fig. 1. An example of different groups of talents.

Qiaozhu Mei et al. [6] builds a transition matrix based on all of the points, and using the properties of the Markov Model, lets $N_T(v)$ be the number of times the path has visited v up to time T, and thus the transition probability $p_T(u,v)$ will be reinforced by; further, they assume $E(N_T) \propto p_v^T$, when the value of T is sufficiently large and the transition matrix tends to be stable; and so they use the value of stationary distribution as the *top-k*.

At the same time, we can also have the understanding that the network must consider the k recommendations which is the same as the Recommender System (RS). The RS needs to recommend k points to represent the network. Y. Koren and Bell [7] won the NetFlix Competition in 2008. Y. Koren [8] proposed some new algorithms which differ from the common Singular Value Decomposition (SVD). SVD can determine the inner information of a matrix but when the matrix data is consistently sparse it cannot be reliably analyzed. Y. Koren [8] put forward the time variable SVD in which they consider that the habits of users are changing with the seasons, and the score of the users is the point where the user-vector and the item-vector intersect. Here, during the competition, Y. Koren [8] used a different combination of the algorithms. The algorithms Y. Koren et al. [7–10] proposed, such as SVD++ and TimeSVD++, are meaningful.

Wei Chen et al. [11] uses a greedy algorithm to get the k points in the Social Network (SN), and these k points have the greatest influence on the SN. Diego Saez-Trumper et al. [12] found the *top-k* trendsetters of the information network. The goals of these researchers are the same as ours, which is to find the k points that best represent the network or have the greatest effect.

3. Neighbor Rank and Gravity Rank

As we mentioned above we know the transition matrix should include time variance. The method TimeSVD++ considers that the prediction varies with time, and after the decomposition of the matrix it is very difficult to explain the meanings of the matrixes that have been decomposed. Also, as the network complexity is increased the required computations required by SVD will increase rapidly.

The *top-k* is a very prominent value currently used in data mining, so we begin seeking our solution through this method. Qiaozhu Mei's method offers better performance in obtaining meaningful results from the experiments; during the computation, they assume $E(N_T) \propto p_v^T$, give the $p_{i,j}^{t+1}$, and analyze the process. Qiaozhu Mei claims that the increase of the N_T will increase the p_v^t, which is a positive correlation; but they do not prove that the formula $p_{i,j}^{t+1}$ can satisfy the equation.

In a given network, G, V is the set of points and E is the set of edges. If we want to use a point to represent the network, then the point must be in the center of the network. Here, we can create the talents network (talents network) via the method in Ref. [1].

Assume the points $x_1, x_2, \cdots, x_n$ are m (Each talents have m research directions) dimensional vectors. We use the cost function to represent the network's cost:

$$f(p) = \min\left\{(p-x_1)^2 + (p-x_2)^2 + \cdots + (p-x_n)^2\right\}$$

Via the derivation:

$$\frac{\partial f(p)}{\partial p} = 0 \Leftrightarrow 2(p-x_1) + 2(p-x_2) + \cdots + 2(p-x_n) = 0$$

We get $p = (x_1 + x_2 + \cdots + x_n)/n$. Commonly, we need the k (2, 3, …, n) points to represent the network. The PageRank which is a