On a patent analysis method for identifying the core technologies of metro in china

Mei Long[1, a *], Tieju Ma[2,b]

[1] East china university of science and technology, Meilong road 130, 200237 Shanghai, China

[2] East china university of science and technology, Meilong road 130, 200237 Shanghai, China

[a] meilongjy@126.com, [b] tjma@ecust.edu.cn

Keywords: Metro. Complex patent network. Node importance contribution. Core technology

Abstract. The purpose of this paper is to identify the core technologies in metro field by analyzing the overall structure of technologies and interactions among of them, which is beneficial to grasp technological trend and advance of metro field in China. For this study, the metro patent data were collected (1986-2015) from patent information analyzing system. And then, we built a complex patent network with their co-occurrence relationship of international patent classification (IPC) information. Considering the node importance depending on its original importance and importance contribution to other nodes and complex patent network's features, we improved the node importance contribution correlation matrix (NICCM) method to calculate node importance in overall network structure.

Introduction

The traffic facility plays a significant role on the development of city, so it is a common phenomenon that there are more automobiles in metropolis, such as Shanghai, Beijing and Guangzhou of China. The automobiles, no doubt, bring much convenience to people's life. However, they should be responsible for a range of social problems as well. Nowadays, people have to be confronted with the challenges of environmental pollution, energy shortage and traffic jam in China. Luckily, many literatures reveal that building metro is an accessible way to solve these problems.

According to statistic, the transport capacity of metro is ten times higher than bus, but it consumes the minimum energy in all transportations except bicycle. Besides, the government always pays great attention to building metro, and it is likely that the total operation mileages of metro will be up to 4500 kilometers in 2050 in china. Therefore, it becomes more important and meaningful to identify core technologies and grasp technological trend and advance of metro by analyzing the overall structure of technologies and interaction among of them [1], which is also the purpose of this study.

Most of the commonly used methods of identifying core technologies are analyzing the complex patent network built with citation information [2,3,4]. However, they have some shortcomings reported in literatures, such as the average 10 years' time-lag between citing-cited patents [5]. Besides, it is difficult to grasp the technological relationship and characteristics with citation analysis, for the citing-cited relatedness is based on the perspective of individual patents [6].

On the contrary, the patent analysis with co-occurrence relationship of international patent classification (IPC) information has some advantages over that. Because a patent is coded with a main IPC and several deputy IPCs, and an IPC indicates a technology. For example, there is a patent A coded as G03B25/00;B61B13/10;G09F19/00, and G03B25/00 is its main IPC, both B61B13/10 and G09F19/00 are deputy IPCs of patent A. Therefore, this study can propose a new approach to build patent network with consideration of the overall interrelationships among technologies.

Centrality measures indicate the position importance of a node in network, and are often used to identify node importance [7]. So they are also helpful for discovering the core technologies through identifying node importance in the patent network. Researchers have tried to propose and improve many measures to identify key nodes including degree centrality, closeness, betweenness , structural holes[7,8].

Since those measures have different emphasis on node importance appraisal for specific issues, and the node importance is affected by multi factors. A multi-attribute decision making method based on the consideration of decreasing the evaluation disparities was proposed to evaluate the node importance in complex network [9]. Meanwhile, considering the time consumption during measure, a semi-local centrality measure was proposed, in which it could realize a tradeoff between the low-relevant degree centrality and other time-consuming measures [10].

Nevertheless, this kind of method takes only into account the interaction between adjacent nodes. But the node importance depends largely on its contribution to both adjacent nodes and non-adjacent nodes. In terms of the node importance contribution, the NICCM method is proposed to identify the core nodes according to evaluating the node importance contribution among different layers in network. In which the adjacent nodes of node are defined as the first layer nodes, and its next adjacent nodes are the second layer nodes etc [7].

While, the NICCM method proposed in an undirected and unweighted network can't work in weighted network, especially the complex patent network. A number of researchers have indicated that the transfer and diffusion of technology requires strong ties [11]. Therefore, there is a necessary attempt on developing the NICCM method to apply it in weighted network. In order to set the relative importance between the number of ties and tie weights, a generalization of shortest distance for weighted network was proposed based on the consideration of making a trade-off between them[12,13]. So it is possible that developing the NICCM method by introducing the general measure of shortest distance in weighted network.

The purpose of this paper is identifying the core technology of metro in china through patent analysis method, which is significant to promote the development of metro industry. In view of the above, we can use the developed NICCM method to realize the goal.

The remainder of this paper is organized as follow. Methodology section introduces the methods of building complex patent network and identifying the core technologies of metro. The experimental results are presented and explained subsequently. Finally, the conclusion and discussion is given.

Methodology
Building complex patent network
As the introduction section showing, we can use the co-occurrence information of patents' IPC to build the complex patent network, which reveals the overall interrelationships among metro technologies. Now we introduce the process of the built complex patent network, meanwhile, give a brief illustration for each step.

Data collection and pre-processing

The patents data of metro area are collected from patent information analyzing system according to relevant search condition, such as [abstract] is metro or [title] is metro. After this stage, the patents' basic attributes are expressed as excel, and we can get the IPC information of each patent. As we can see from Fig. 1, from 1986 to 2013, for most of metro patents (classified as B, E and H), the number of patents' increase, but it begins to decrease since 2013, this is because the time-lag between applying-publishing patent. But the patents' IPC co-occurrence matrix can't be constructed directly based on these original data, which need to be pre-processed and transformed to the structured document (e.g. patent name, main IPC, deputy IPC1, and deputy IPC2 etc).
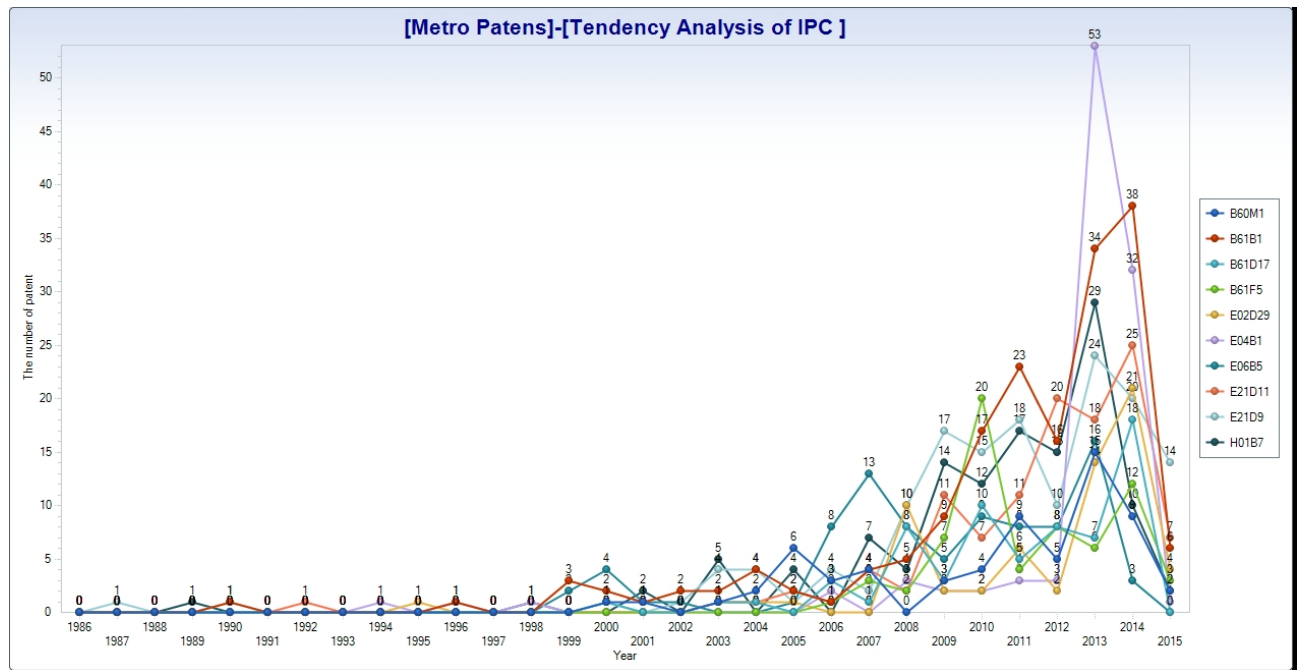
Fig. 1 Development tendency of IPC

Patents' IPC co-occurrence matrix

For a patent is coded with different IPC, and the same IPC may occur in different patent, the number of co-occurrence can be figured out between two arbitrary IPCs. Therefore, the patents' IPC co-occurrence matrix of metro, donated $T = [t_{ij}]$, is constructed and exemplified partly in Table 1. In this table, there are 3318 IPCs. t_i means the ith IPC, and the number of patents' IPC co-occurrence is arranged by decimal value: $t_{1,2}=62$ means that IPC t_1 and t_2 occur together 62 times in same patent, while '0' means the two IPCs do not.

Table 1 Example of patents' IPC co-occurrence matrix

Patent IPC	t_1	t_2	...	t_{3317}	t_{3318}
t_1	0	62	...	0	12
t_2	62	0	...	0	9
...	...	...	0	...	...
t_{3317}	0	0	...	0	0
t_{3318}	12	9	...	0	0

Building complex patent network

The complex patent network is built on the basis of relevant relationships represented in patents' IPC co-occurrence matrix. Accordingly, the network contains 2811 nodes, and if t_{ij} isn't equal to zero, there will be a directed edge between node i and j, otherwise the two nodes can't be connected. That is, as $t_{ij} \neq 0$ if there is an edge between node i and j, $t_{ij} = 0$ otherwise. Besides, when t_{ij} isn't equal to zero, the value of t_{ij} is the tie weight between node i and j. Therefore, as the Fig. 2 presenting, the complex patent network can be constructed based on the co-occurrence relationship among IPCs.
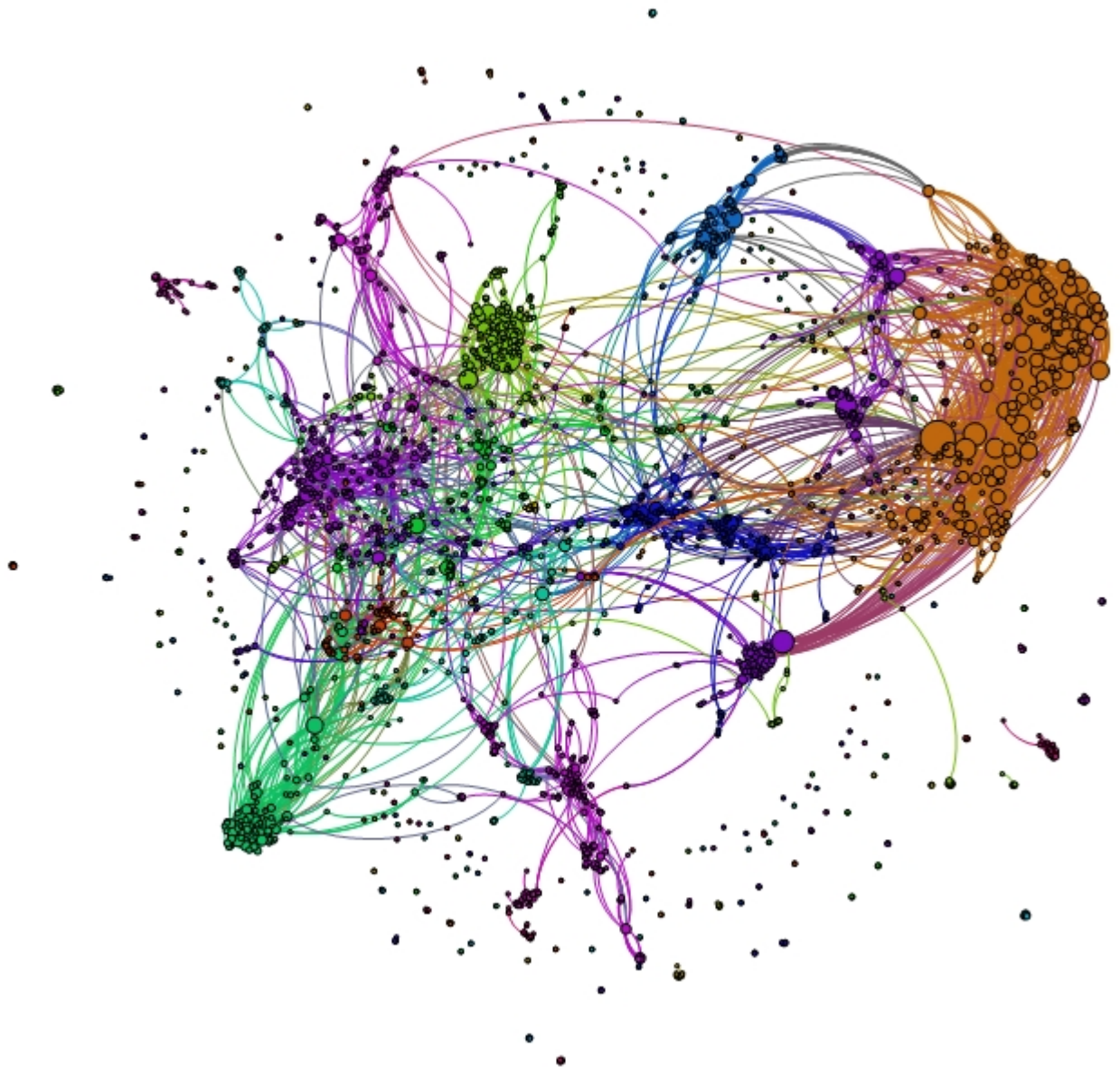
Fig. 2 The complex patent network

Identifying the core technologies

The node importance depends on its original importance and importance contribution to other nodes, in the weighted complex patent network, node i contributes its initial importance to node j, and also accept the contribution coming from node j at the same time. Then how to describe the complex contribution relationship becomes a crucial issue. Next, we will introduce the new NICCM method-building process.

  Shortest distance in weighted network

  Researchers have tried to identify the shortest path in weighted network, most of such attempt defined the weights as costs of information transmitting since high values indicated weak or costly ties. While the tie weights were also referred to strength not the cost of ties, and the tie weights needed to be reversed before using [13,14].

  Node efficiency

  In a complex network, the node plays a very important role during information transferring, which reveals the contribution to network efficiency[15]. So the node efficiency can be defined as below:

$$E_i = \frac{1}{n}\sum_{j=1,j\neq i}^{n}\frac{1}{d^W(i,j)}.$$

(1)

Where E_i is the efficiency of node i, $d^W(i,j)$ is the shortest distance between node i and j, and n is the number of all nodes in the complex patent network.

Node importance contribution correlation matrix

In an undirected and unweighted network, the importance correlation parameter presents the contribution probability from node i to node j, and strength correlation parameter expresses the dependence strength of node j to i[16]. As the introduction section saying, we can develop the NICCM method to define node importance by introducing the general measure of shortest distance in weighted network. So we will get a new node importance contribution correlation matrix (NNICCM) as follow:

$$I = \begin{bmatrix} 0 & \frac{1}{d^W(1,2)}\mathcal{P}_{1,2}E_1 & \cdots & \frac{1}{d^W(1,n)}\mathcal{P}_{1,n}E_1 \\ \frac{1}{d^W(2,1)}\mathcal{P}_{2,1}E_2 & 0 & \cdots & \frac{1}{d^W(2,n)}\mathcal{P}_{2,n}E_2 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{d^W(n,1)}\mathcal{P}_{n,1}E_n & \frac{1}{d^W(n,2)}\mathcal{P}_{n,2}E_n & \cdots & 0 \end{bmatrix}.$$

(2)

For matrix I, E_i is the base efficiency of node i, which can contribute $\frac{1}{d^W(i,j)}\mathcal{P}_{i,j}$ of its original importance to other nodes (adjacent nodes and non-adjacent nodes), where $\frac{1}{d^W(i,j)}$ is the strength correlation parameter, used to describe the impact of distance on the dependence strength, and $\mathcal{P}_{i,j}$ is the importance correlation parameter expressing the contribution probability from node i to node j [16].

Finally, we can define the node importance I_j of node j as

$$I_j = E_j * \sum_{i=1,i\neq j}^{n}\frac{1}{d^W(i,j)}\mathcal{P}_{i,j}E_i.$$

(3)

Results

In view of the above, the node importance in complex patent network can be calculated by using the improved NNICCM method, and we can at last grasp the core nodes according to the node importance rankings. Meanwhile, the core technologies of metro can be identified, and then we can analyze the metro technological development trend and advance of metro by comparing the overall technologies ranking. Fig. 3 shows the growth tendency of metro patents, in which we can draw a conclusion that the metro industry is rapidly developing in general. Furthermore, we can also forecast the development tendency of metro's core technologies in the future.
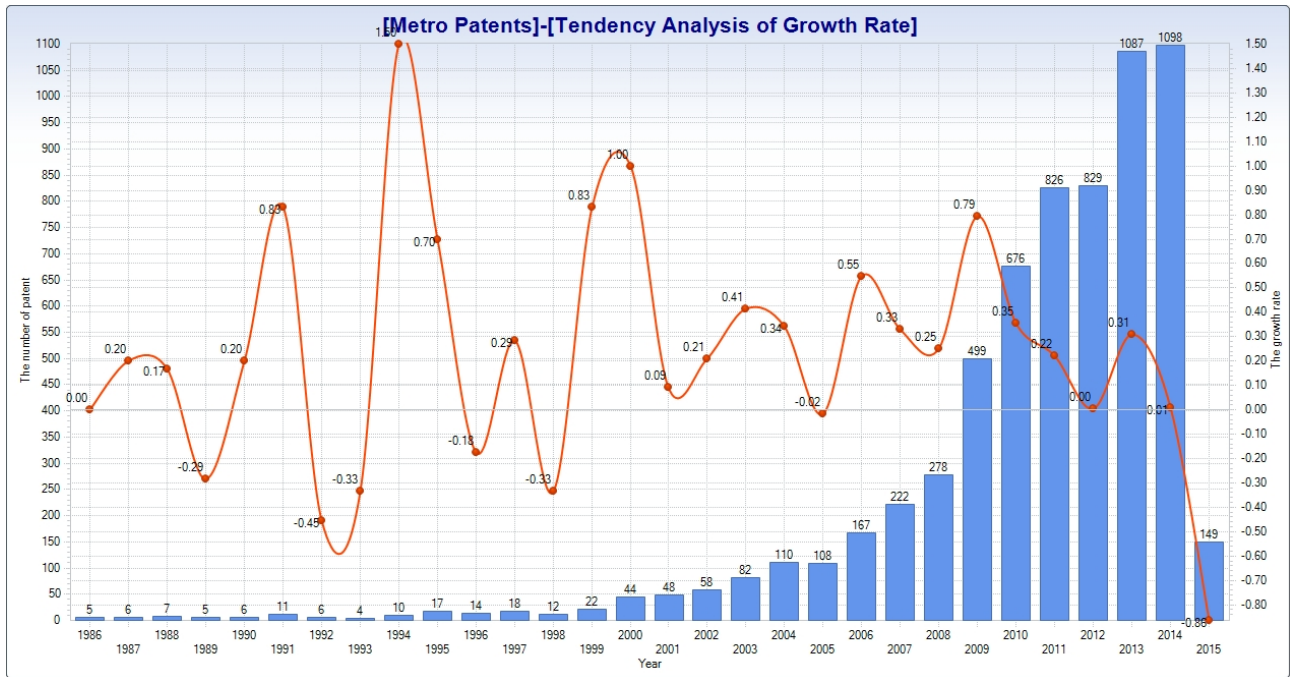
Fig. 3 The complex patent network

Conclusion and discussion

This study was motivated by the need of resolving some urgent social problems in China, for instance, environment pollution, energy shortage and traffic jam, and many literatures reveals that developing metro is an efficient way to alleviate them. So the strategic importance of analyzing current development status and grasping the technology opportunities has become more apparent.

Firstly, this study proposed a new approach to build complex patent network in terms of interrelationship among metro technologies by creating the IPC co-occurrence matrix. Secondly, for analyzing node importance in the complex patent network, this study improved NICCM method and calculated the node contributions in global network structure. In this method, the technology importance was evaluated in overall technology structure, which was more feasible and valid than traditional method to identify core technologies. Finally, based on the identified core metro technologies, we could forecast their development tendency according to combine with technology life cycle theory.

In our future work, we will study the underlying technological structure features of metro by comparing with other developed countries, which is beneficial to deeply grope for the new development direction of metro.

Acknowledgments

References

[1] Lee H, Kim C, Cho H, et al. An ANP-based technology network for identification of core technologies: A case of telecommunication technologies [J]. Expert Systems with Applications, 2009, 36(1): 894-908.

[2] Duch-Brown N, Costa-Campi M T. The diffusion of patented oil and gas technology with environmental uses: A forward patent citation analysis[J]. Energy Policy, 2015, 83: 267-276.

[3] Newman M E J. Modularity and community structure in networks [J]. Proceedings of the National Academy of Sciences, 2006, 103(23): 8577-8582.

[4] Rodriguez A, Kim B, Lee J M, et al. Graph kernel based measure for evaluating the influence of patents in a patent citation network[J]. Expert Systems with Applications, 2015, 42(3): 1479-1486.

[5] Kim C, Seol H. On a patent analysis method for identifying core technologies [J].Intelligent decision technologies. Springer Berlin Heidelberg, 2012: 441-448.

[6] Yoon B, Park Y. A text-mining-based patent network: Analytical tool for high-technology trend [J]. The Journal of High Technology Management Research, 2004, 15(1): 37-50.

[7] Burt R S. Structural holes and good ideas [J]. American journal of sociology, 2004, 110(2): 349-399.

Hu P, Fan W, Mei S. Identifying node importance in complex networks [J]. Physica A: Statistical Mechanics and its Applications, 2015, 429: 169-176.

[8] Zhang X, Zhu J, Wang Q, et al. Identifying influential nodes in complex networks with community structure[J]. Knowledge-Based Systems, 2013, 42: 74-84.

[9] Hui Y, Zun L, Yong-Jun L. Key nodes in complex networks identified by multi-attribute decision-making method[J]. 2013 (In Chinese).

[10] Chen D, Lü L, Shang M S, et al. Identifying influential nodes in complex networks[J]. Physica a: Statistical mechanics and its applications, 2012, 391(4): 1777-1787.

[11] Tsai W. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance [J]. Academy of management journal, 2001, 44(5): 996-1004.

[12] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths [J]. Social Networks, 2010, 32(3): 245-251.

[13] Newman M E J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality [J]. Physical review E, 2001, 64(1): 016132.

[14] Brandes U. A faster algorithm for betweenness centrality [J]. Journal of Mathematical Sociology, 2001, 25(2): 163-177.

[15] Xuan Z, Feng-Ming Z, Ke-Wu L, et al. Finding vital node by node importance evaluation matrix in complex networks[J]. 2012.

[16] Hu P, Fan W, Mei S. Identifying node importance in complex networks [J]. Physica A: Statistical Mechanics and its Applications, 2015, 429: 169-176.