Evolution of patent citation network for hybrid rice breeding

Tian-Hua Song [1,2, a *] and Qing-Pu Zhang [2, b]

[1] Library Harbin Institute of Technology, 73, Huanghe Road, Harbin 150090, China.

[2] School of Management Harbin Institute of Technology, Harbin 150001, China

[a]tianhua_s@163.com,[b]zzqp2000@126.com

Keywords: Patent Citation Network; Evolution; Hybrid Rice Breeding

Abstract. This study models the evolution of technology innovation network. Results argue that the growth of hybrid rice innovation network fellows S-curve by time; the innovation network fits preferential attachment as amended aging Lognormal distributions. And it implies that the degree distribution of the technology innovation network follows a power law distribution.

Introduction

Innovation plays the most important role in economic development. Patents and patent citations are two valid indexes valuing technological innovation and diffusion. And patent citation networks can represent well technological innovation networks patented. Csárdi et al. explored dynamic of global patent citation network based on U.S. Patent and Trademark Office' (USTPO)patent data and argue attractiveness function A (k, l) determining the likelihood that a patent will be cited shows power law aging and preferential attachment [1]. Valverde et al. implied the technology innovation evolution model including Weibull form aging and preferential attachment [2].Other studies explored technological innovation development of worldwide nanotechnology [3], emerging [4] and developing trend [5-6] of some technologies. Based on technological trajectories some scholars have performed studies on technological diffusion [7-8].

But this study explores its connecting probability model of patent citation networks based on aging distribution and preferential attachment. And based on the evolution dynamic mechanism of the technology innovation network, we model the evolution of the technology innovation network for hybrid rice breeding and use mean-field to calculate accurately the degree distribution of the technological innovation network.

Research method and model

The data were obtained from patent data in the Derwent Innovation Systems. They were acquired in January 2011 which were retrieved using a topic search query "hybrid rice" as the subject, and covering the time span from 1976 to 2010, 221 patents of hybrid rice were firstly retrieved. Then 84 of their patent citations including forward citations and backward citations were obtained. In this patent-citation network, each node is a patent and each edge represents a citing or cited relationship. The out-degree of each node represents the number of citations of a patent, while the in-degree represents the number of references to other patents. For example, $A \leftarrow B$ means that A is cited by B or B references A, representing technological diffusing direction from A to B. Then matrixes of networks based node (patent and patent citation) and edge (citing or cited relationship) are constructed by cumulative time (for example from 1976 to 1977, from 1976 to1978,… from 1976 to 2010) to examine evolution of the network. The size of the biggest hybrid-rice patent citation network from 1976 to 2010 is 116 nodes and 169 edges (outliers ignored).

Degree distribution. The degree of node i, ki, is defined as the number of other nodes connecting to node i. Counting the number of nodes having the same degree yields the degree distribution function, denoted as P(k). Different networks have different degree distribution functions

Growth rate. We define that number of new node every time step joins network is the growth rate of network. Usually we use number of new nodes every year entering network as growth rate of network.

Average distance evolution. The average distance is defined as the average number of edges which connect to form the shortest path between any two nodes in the networks. Average distance evolution can indicate how "the distance of network" changes by time. A short average path length means that technology will move to different parts of the network more quickly.

Clustering coefficient evolution. The clustering coefficient is an important parameter measuring the degree of network clustering. The clustering coefficient of node i, denoted as $C_i \in [0, 1]$, is the ratio of the actual number of edges to the possible number of edges between node i and other nodes in networks. Average clustering coefficient is $1/N \cdot \Sigma C_i$, representing the average degree of network clustering.

We define the computing model framework that we will use for studying technological diffusion network. Under preferential attachment mechanism new elements entering the system connect with an old node with a probability $\Pi(k_i)$ that is proportional to its degree k_i.

So we assume the connectivity probability kernel function of an old patent cited by new patents in technological innovation networks, including preferential attachment proportional to its degree k and connectivity proportional to its age. The connectivity probability kernel function is

$$\prod Ker(k_i, A) = \Omega(f(k_i), y(A)) \tag{1}$$

The $\prod Ker(k_i, A)$ function determines the evolution of the network.

Results and Discussion

The degree distribution of the hybrid-rice patent citation network and its fitted function are shown in Fig. 1.The degree distribution trend function of the patent citation network for hybrid rice is

$$P(k) = a'k^{-g} \tag{2}$$

It is a power law distribution; with power exponent is -2.01 and a' is 81.12. So the technology innovation network is a scale-free network.
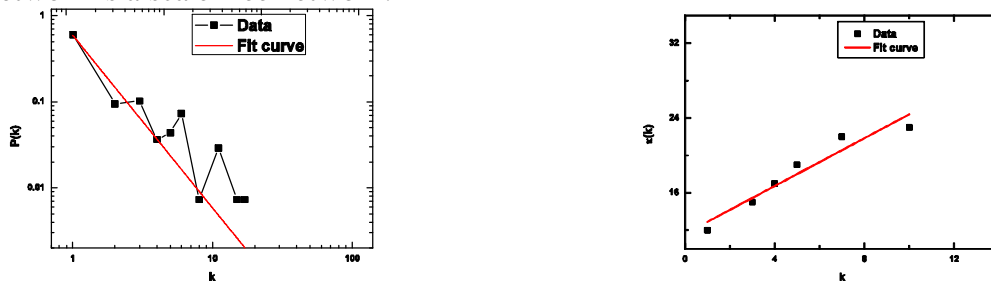

Fig.1(left) Degree distribution of hybrid-rice patent network, where P(k) is degree distribution,k is degree.

The important mechanisms of scale-free network are preferential attachment and grow. We first examine preferential attachment. For sample data in one year is very small, so we extract the patent citation data from 1963 to 1994 as degree k and the patent citation data from 1995 to 2001 as attachment probability $\prod(k_i)$. To reduce the fluctuations on the data from the network, we calculated the cumulative preferential attachment, defined as [9]

$$k(k) = \sum_{k_i=0}^{k} \prod(k_i) \tag{3}$$

Then we examine the attachment of patent citation from 9 to 11 ages as enough samples. Result shows a well linear preferential attachment in Fig. 2. It implies that age of cited patent has important effect on the network attachment. So at next step we will examine aging attachment.

Fig.2(rihgt). Preferential attachment of technology innovation network for hybrid rice based on age-modified, the data, cumulative preferential attachment is the cumulative data from 9 to 11 ages proportional to its degree k.

Aging distribution of hybrid rice breeding innovation network follows LogNormal distribution different from the distributions in references [1-2, 10-11]. Aging distribution of hybrid rice breeding innovation network is

$$\prod(A) = y_0 + \frac{D}{\sqrt{2p}\,wA} e^{-\left[\frac{\ln\frac{A}{A_c}}{2w^2}\right]^2} \tag{4}$$

Where f(A) represents patent citation number by age, y0 is the primary connectivity probability(patent citation number) of the technology innovation network, Ac is the peak age, D and w are two parameters of system; y0=−1.40, Ac = 4.93, D = 43.89 and w=0.46 in the network. It implies that its decay by age is slower than the global technological innovation networks (see in Fig. 3).

Fig. 3(left). Aging decay distribution of hybrid rice breeding innovation network

Fig. 4(right). The cumulative growth rate of the technological innovation networks for hybrid rice breeding

Because growth rate of the network is fluctuations on the data, so we calculated the cumulative growth rate. The best fitted curve is an S-curve named Sigmoidal Logistic function and not a linear curve shown in Fig. 4. Its distribution function is

$$Y_G(t_i) = \frac{a}{1 + e^{-b(t_i - t_c)}} \tag{5}$$

where   is the rate of nodes entering the network at its saturation stage,   is a parameter of system controlling the curve shape; tc is the inflection time point. New joining innovation increases continually at primary stage and gets the top at the inflection time point. After the inflection point the growing rate will decreased slowly, and finally tend to constant when the technology innovation gets to the technology saturation stage. Where   =127.8,   =0.21 and tc=15.03. From Equation (3) and Equation (6), we can get a composite connectivity probability function $\prod Ker(k_i, A)$ for the technology innovation network of hybrid rice breeding, which is the product of a preferential attachment function $\frac{k_i}{\Sigma k_j}$ and an aging function following Lognormal distributions by Equation (10)，where β is a tuning parameter.

$$\prod Ker(k_i, A) = \frac{k_i}{\sum k_j} \cdot b \left(y_0 + \frac{D}{\sqrt{2p}\,wA} e^{-\left[\frac{\ln\frac{A}{A_c}}{2w^2}\right]^2} \right) \tag{6}$$

Modeling technology innovation network for hybrid rice

We use the obtained numerical results to construct a simple model for the evolution of the technology innovation network. It is important to emphasize that the purpose of the model is to capture the main mechanisms that affect the evolution of the innovation network, and not to incorporate every numerical detail of the measured web. It is important to emphasize that the properties of the technology innovation network for hybrid rice are not unique too. Though normally most technology innovation domains have themselves' age distribution for their technology life circle, most of their connectivity

probability is product of a preferential attachment function $\frac{k_i}{\Sigma k_j}$ and an aging function. The WWW is also a complex evolving network, where links are added (and removed) changing as a linked page' age, the network topology being profoundly determined by these dynamical features.

We denote by ki(t) the number of links node i has at time t; by T(t) and N(t) the total number of links and total number of nodes at time t, respectively. We assume that all nodes present in the system are more active, i.e. they can make further innovation. In agreement with figure 3, we consider that new innovation joins the field at a linear rate, leading to

$$N(t_i) = \frac{a}{1 + e^{-b(t_i - t_c)}}$$
(11)

The average number of links per node in the system at time t is thus given by:

$$\langle k \rangle = \frac{T(t)}{N(t)}$$
(12)

A power law distribution for the technology innovation network also indicates that new nodes link to the existing nodes with preferential attachment. Thus, if node i has ki links, the probability that an incoming node will connect to it is given by Equation (1).

We have thus formulated the dynamical rules that govern our evolving network model, capturing the basic mechanism governing the evolution of the technology innovation network for hybrid rice breeding:

Nodes join the network at a liner rate by equation (8). And emphatically every technology innovation network can have normally itself growing mode.

Incoming nodes link to the already present nodes following preferential attachment proportional to its degree by Equation (1).

Aging attachment distribution of hybrid rice breeding innovation network is Lognormal distribution by Equation (5).

A composite connectivity probability function $\prod Ker(k_i, A)$ for the patent citation network of hybrid rice breeding is Equation (7).

The front two features of the model should be mostly same as many technological innovation networks and other webs including www web, science papers citation networks and transport networks. The back two features lie in www web and science papers citation networks. So the modeling is general in some domains.

Conclusions

Technological innovation network for hybrid rice breeding is a scale-free network fitted with an allometric power law distribution with exponent -2. That the network growing by Sigmoidal Logistic function by time implies that innovation of hybrid rice has varied activity. That the probability new nodes connecting with old nodes in the networks follows a composite connectivity probably model as preferential attachment proportional to its degree and aging Lognormal distributions.

Studies in future may explore more key topology details and text mining of technology innovation network.

Acknowledgment

References

[1] G. Csárdi, K. J. Strandburg, L. Zalányi, J. Tobochnik, P. Érdi. Modeling innovation by a kinetic description of the patent citation system .Phys. A.374 (2007)783–793

[2] S. Valverde, R. V. Solé, M. A. Bedau, N. Packard. Topology and evolution of technology innovation networks. Phys. Rev. E.76 (2007)056118

[3] X. Li, Y.L. Lin, H.C Chen, M. C. Roco. Worldwide nanotechnology development: a comparative study of USPTO, EPO, and JPO patents (1976–2004). J Nanopart Res .9 (2007)977–1002

[4] P.L. Chang, C.C. Wu, H. J. Leu .Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. *Scientometrics*, 82 (2010) 5–19

[5] S.B. Chang, K.K. Lai, S.M. Chang. Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technological* . Technol. Forecast. Soc. Change 76 (2008) 107–117.

[6] J.Yoon, K. Kim. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics* 88 (2011)213–228

[7] R. Fontana, N. Alessandro, B. Verspagen. Mapping technological trajectories as patent citation networks: an application to data communication standards. *SPRU* [Science & Technology Policy Research: University of Sussex] *Electronic Working Series*, (2008)Paper 166, 2

[8] B. Verspagen. Mapping Technological Trajectories as Patent Citation Networks. A Study on the History of Fuel Cell Research. (2005)Working Paper 05.11

[9] H. Jeong, Z. Néda, A.-L. Barabási, Measuring preferential attachment for evolving networks. Euro. Phys. Lett. 61(2003) 567

[10] A.-L. Barabàsi, R. Albert, Emergence of scaling in random networks. Sci. 286 (1999)509-512.

[11]A.-L. Barabàsi, R. Albert, H. Jeong, Mean-field theory for scale-free random networks. Phys. A 272 (1999)173-187.