Two-dimensional Technology Profiling of Patent Portfolio

C.-H. Kuan[1,3], W.-M. Tu[1], D.-Z. Chen[2,3]

[1]Graduate Institute of Patent, National Taiwan University of Science and Technology, Taipei, Taiwan R.O.C.

[2]Dept. of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, Taiwan R.O.C.

[3]Center for Research in Econometric Theory and Applications, National Taiwan University, Taipei, Taiwan R.O.C.

(maxkuan@mail.ntust.edu.tw, patent.ntust.benton@gmail.com, dzchen@ntu.edu.tw)

Abstract – **Noticing that some technology content is not reflected by patents' individual classification symbols but by the co-assignment of these symbols, this study proposes to represent the technology content of a patent portfolio using a two-dimensional matrix, referred to as the *profile matrix* of the portfolio. The element M*ij* of a profile matrix M counts the co-assignment frequency of symbols C*i* and C*j*, and the element M*ii* is the individual assignment frequency of symbol C*i*. The profile matrix not only covers the traditional one-dimensional patent classification analysis, but also provides a more comprehensive picture to the portfolio's technologies. The profile matrix may be applied to detect the similarity or relatedness between patent portfolios, monitor the shift of an entity's R&D direction, and discover the emergence of new, cross-disciplinary technology.**

Keywords – **Patent classification, Classification symbols, Co-assignment, Profile matrix.**

I. INTRODUCTION

Every patent is associated with one or more classification symbols. These symbols are a valuable source of information as, in addition to being readily available from the patent's bibliographic data, they are assigned by a professional examiner during a rigorous patentability determination process by categorizing the patent's technology content according to a standardized classification scheme.

All classification schemes provide a tree-like hierarchical taxonomy of technologies. Currently, most countries adopt International Patent Classification (IPC), which is maintained by World Intellectual Property Organization (WIPO) and offers about 70,000 technology areas [1]. From mid-2015, European Patent Office (EPO) and U.S. Patent and Trademark Office (USPTO) started to use their jointly developed Cooperative Patent Classification (CPC) [2], which provides more than 260,000 technology areas. Due to the endorsement of EPO and USPTO, two of the largest patent offices, it is expected that CPC would eventually be adopted by all patent offices across the globe.

For example, Fig. 1 is a screen capture of U.S. Patent No. 10,061,509, titled "Keypad control," from USPTO full-text database. As illustrated, the patent is associated with six CPC symbols listed under "Current CPC Class", in addition to four IPC symbols under "Current International Class" (so that the patent is still searchable using IPC by other patent offices), and a dummy U.S. classification symbol 1/1 under "Current U.S. Class" (as the U.S. self-developed scheme is discarded).

Using one of the CPC symbols, G06F3/017, as an example, its association with the patent indicates that the patent involves an area of technologies "Gesture based interaction," which is at the 6th level of CPC's hierarchical taxonomy subordinated to the higher and more abstract levels outlined in Table I.

Various types of patent classification analysis have been proposed. The following are some examples. The number of different classification symbols assigned to the patents of an entity (e.g., a company, an institute, a country, a technology field) are deemed to reflect the entity's technology diversity [3][4]. If two classification symbols have high co-assignment frequency to patents, the technical areas denoted by the classification symbols are considered to be more related [5][6][7]. The classification symbols assigned to patent portfolios are compared to detect their technology similarity [8][9]. The classification symbols of a patent's forward and backward citations are used to evaluate the patent's "generality" and "originality" [10]. The similarity between patents may be obtained based on classification symbols' structural information in the hierarchical classification scheme using the so-called Jaffe Distance [11]. Classification symbols may also be used to see how firms diversify their innovative activities across technology fields [12].

The most common patent classification analysis, which is available from all patent analysis systems known to the authors, is to investigate the R&D focus of an entity by the assignment frequencies of the classification symbols of its patent portfolio, usually manifested in a bar chart. Fig. 2 shows the result of such an analysis to a company's 2013 patent portfolio whose symbols are reduced to the 4th or main-group level. Based on the diagram, the company is considered to have its R&D focused mainly in the field denoted by the most frequently assigned symbol G06F3, which is related to various technologies about data input into or output from computers (see Table I).

Fig. 2 is similar to a signal's frequency spectrogram. Just like a frequency spectrogram reflects the signal's characteristics, Fig. 2 technologically profiles the company's patent portfolio. One may imagine that each entity has its own technology profile graphically

This work was financially supported by the Center for Research in Econometric Theory and Applications (Grant no. 107L900204) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, and by the Ministry of Science and Technology (MOST), Taiwan, under Grant No. MOST 107-3017-F-002-004-.

described by a diagram like Fig. 2, which provides a comprehensive picture about an entity' distribution and focus of technologies.

II. METHODOLOGY

The one-dimensional technology profile shown in Fig. 2 has a number of shortcomings. Firstly, a patent often has multiple classification symbols and each captures only a portion of the patent's technology content. This study notices that there may be some "hidden" content not reflected by individual symbols, but by the joint assignment, or *co-assignment*, of two or more symbols. In other words, for a patent having two classification symbols {C*i*, C*j*}, the patent's technology content may involve those reflected by individual symbols {C*i*} and {C*j*}, and those by the co-assignment of {C*i*} and {C*j*}.

Taking the patent US7,657,849 as an example. It is Apple's first patent disclosing the slide-to-unlock function on all iPhone and iPad devices. The patent has ten CPC symbols involving four 4th-level symbols listed in Table II.

Current U.S. Class:		**1/1**
Current CPC Class:		G06F 3/04886 (20130101); G06F 3/013 (20130101); G06F 3/0304 (20130101); G06F 3/044 (20130101); G06F 3/017 (20130101); G06F 2203/04804 (20130101)
Current International Class:		G06F 3/0488 (20130101); G06F 3/01 (20060101); G06F 3/044 (20060101); G06F 3/03 (20060101)

Fig. 1. Various classification symbols from patent US10,061,509.

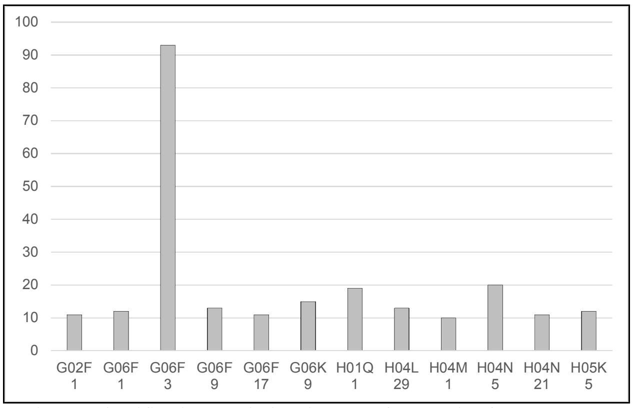

Fig. 2. Classification symbol assignment frequencies for a company.

TABLE I
HIERARCHICAL STRUCTURE OF CPC SYMBOL G06F3/017

Level	Symbol	Title
5th or sub-group level	G06F3/01	Gesture based interaction, e.g. based on a set of recognized hand gestures
4th or main-group level	G06F3 or G06F3/00	Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements
3rd or sub-class level	G06F	Electric digital data processing
2nd or class level	G06	Computing; Calculating; Counting
1st or section level	G	Physics

TABLE II
MEANING OF US7,657,849 CLASSIFICATION SYMBOLS

Symbol	Title
G06F3	Input arrangements for transferring data to be processed into a form capable of being handled by the computer; Output arrangements for transferring data from processing unit to output unit, e.g. interface arrangements
G06F21	Security arrangements for protecting computers, components thereof, programs or data against unauthorised activity
H04M1	Substation equipment, e.g. for use by subscribers; Analogous equipment at exchanges
H04M 2250	Details of telephonic subscriber devices

According to their titles under the CPC scheme, one may see that G06F3 (related to data input/output) and G06F21 (related to authentication) individually and separately capture a part of the slide-to-unlock function. The gist of slide-to-unlock seems to be more appropriately reflected by combining G06F3 and G06F21 (data input/output for authentication).

A second shortcoming of the one-dimensional technology profile is that it may fail to differentiate some patent portfolios. For a simplified example, a patent portfolio P includes two patents with classification symbols {C_A, C_D, C_F}, {C_B, C_D, C_F}, and another portfolio Q includes three patents with classification symbols {C_A, C_D}, {C_B, C_F}, {C_D, C_F}. The one-dimensional technology profile counts the symbols individually and separately, and both portfolios show an identical distribution of symbol assignment frequencies as {C_A, C_B, C_C, C_D, C_E, C_F}={1, 1, 0, 2, 0, 2}.

This study proposes a two-dimension matrix representation of a patent portfolio's classification symbols, referred to the *profile matrix* of the patent portfolio. A patent portfolio is said to have N classification symbols {C_1, C_2, …, C_N}, and its profile matrix M is a NxN square and symmetric matrix where M*mn*=M*nm* is the frequency of co-assignment of classification symbols C*n* and C*m*, and M*nn* or M*mm* is the frequency of assignment of classification symbol C*n* or C*m*, to the patents of the portfolio.

Then the previous exemplary portfolios P and Q have respective profile matrices shown in Fig. 3, where, for simplicity's sake, they are represented as upper triangular matrices and zero frequency is left blank. Even though the portfolios P and Q have identical one-dimensional technology profile {1, 1, 0, 2, 0, 2}, the difference between the two portfolios is obvious from their two-dimensional profile matrices. Please note that the one-dimensional technology profile {1, 1, 0, 2, 0, 2} is included by the profile matrices along their diagonals.

	A	B	C	D	E	F
A	1			1		1
B		1		1		1
C						
D				2		2
E						
F						2

	A	B	C	D	E	F
A	1			1		
B		1				1
C						
D				2		1
E						
F						2

Fig. 3. Profile matrices for fictitious portfolios P (left) and Q (right).

Some issues about the profile matrix are described as follows. Due to the numerousness of classification symbols, it is a common practice to reduce classification symbols to a higher, such as 3rd or 4th, level so that the profile matrix has a smaller dimension and may be processed more efficiently. As such, a patent may have duplicated 3rd- or 4th-level classification symbols. For example, the patent of Fig. 1 has several G06F3 after the symbols are reduced to the 4th level. These duplicated symbols are handled as different symbols. Say a patent has symbols {C*i*, C*i*, C*k*} after reduction. Then the elements M*ik* and M*ki* of the profile matrix would be two as the symbols {C*i*} and {C*k*} are co-assigned twice to this patent. Similarly, the element M*ii* would be two.

Another issue is that, according to CPC scheme, a patent's classification symbols include *invention* classification symbols covering the novel and non-obvious information contained in the patent, and *additional* classification symbols involving other information considered helpful for searching [1][2]. This study chooses to use both the invention and additional symbols in constructing the profile matrix. However, whether to use only the invention symbols or both is still an open question and the effectiveness of the two approaches should be investigated in the future.

III. RESULTS

The profile matrix, as it captures more detailed information about a portfolio's technology content, may be utilized to reveal how an entity shifts its technology development direction over time in a more accurate manner.

To verify this application, this study chooses a world well-know, Taiwan-based smart phone manufacturer, HTC Corporation, for empirical observation. The company's 962 U.S. utility patents issued from 2012 to 2016 and their CPC symbols (including both invention and additional classification symbols) are collected. The profile matrices for patents issued in the respective years are constructed year by year using 4th-level CPC symbols. Multidimensional scaling (MDS) [13] is then applied to the profile matrices so as to visually detect the variation in the CPC symbols' co-assignment frequencies over the years. In the MDS diagrams, symbols having shorter distances are more frequently co-assigned. For brevity's sake, only the profile matrices for the years 2012, 2014, and 2016 are shown in Figs. 4, 5, and 6. Due to the rather large dimensions of these matrices, only portions of them including symbols of higher co-assignment frequencies are provided. The corresponding MDS diagrams are shown in Figs. 7, 8, and 9, respectively.

	H04L43	H04L27	H04N21	H04L5	H04M1	H04L1	H04L12	H04H20
H04L43	1014	1014	624	546	0	390	390	390
H04L27	1014	1014	624	546	0	390	390	390
H04N21	624	624	393	336	0	240	240	240
H04L5	546	546	336	294	0	210	210	210
H04M1	0	0	0	0	216	0	6	0
H04L1	390	390	240	210	0	180	150	150
H04L12	390	390	240	210	6	150	161	150
H04H20	390	390	240	210	0	150	150	150

Fig. 4. A partial profile matrix for HTC's 2012 U.S. patents.

	H04M1	H04L43	H04L27	G06F3	G06F1	H04L5	H01L2224	H04L1
H04M1	273	0	0	39	71	0	0	0
H04L43	0	171	169	0	0	91	0	65
H04L27	0	169	169	0	0	91	0	65
G06F3	39	0	0	169	12	0	0	0
G06F1	71	0	0	12	125	0	0	0
H04L5	0	91	91	0	0	109	0	53
H01L2224	0	0	0	0	0	0	100	0
H04L1	0	65	65	0	0	53	0	95

Fig. 5. A partial profile matrix for HTC's 2014 U.S. patents.

	H04M1	G06F3	H04N5	G06F1	H04W76	G06F17	H04L9	G06T2207
H04M1	220	26	10	62	0	0	0	0
G06F3	26	207	7	55	0	16	0	3
H04N5	10	7	185	8	0	0	0	28
G06F1	62	55	8	151	0	0	0	0
H04W76	0	0	0	0	98	0	0	0
G06F17	0	16	0	0	0	77	0	0
H04L9	0	0	0	0	0	0	75	0
G06T2207	0	3	28	0	0	0	0	66

Fig. 6. A partial profile matrix for HTC's 2016 U.S. patents.

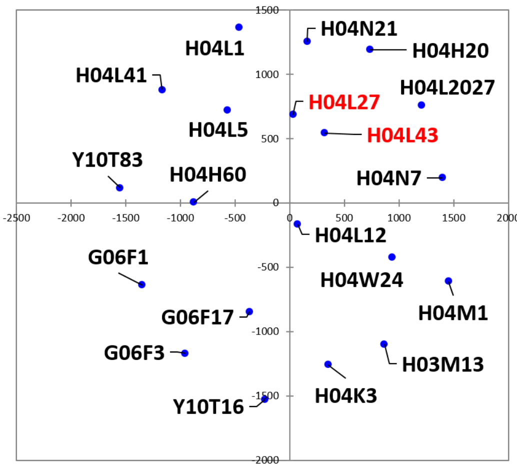

Fig. 7. MDS diagram for HTC's 2012 profile matrix.

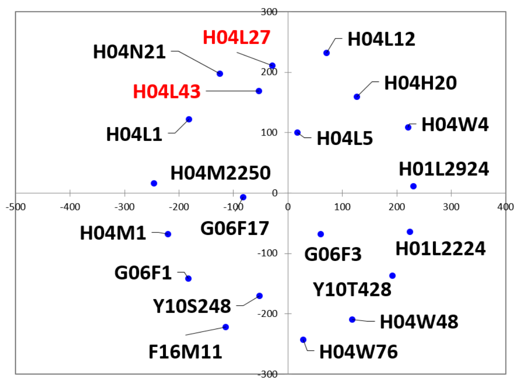

Fig. 8. MDS diagram for HTC's 2014 profile matrix.

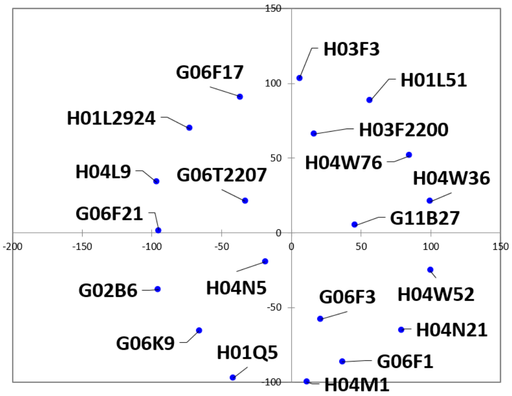

Fig. 9. MDS diagram for HTC's 2016 profile matrix.

These profile matrices and MDS diagrams reveal that HTC's patents in the year 2012 encompass identical technology content, and the co-assignment of classification symbols mainly occurs among symbols H04L1, H04L5, H04L12, H04H20, H04N21, H04L27, H04L43. Among them, H04L27 and H04L43 have the highest co-assignment frequency and this is reflected by their shortest distance in the MDS diagram of Fig. 7, where the two symbols are specifically marked in red.

According to the CPC scheme, H04L27 is about "Modulated-carrier systems," and H04L43 is about "Arrangements for monitoring or testing packet switching networks," both under "Transmission of digital information" (H04L). From the titles of some of these patents, such as "Systems and methods for orthogonal frequency divisional multiplexing" and "System for multiple use subchannels," HTC seems to focus more on developing advanced communication capability for its mobile phone products in 2012.

During the year 2014, even though H04L27 and H04L43 still have the highest co-assignment frequency, some pairs of symbols show more obvious co-assignment frequencies, whereas those significant from the previous years show declined frequencies. This phenomenon indicates that HTC has shifted its R&D direction towards new areas. For example, G06F1 and G06F3 are both "data processing" related symbols. Their more frequent assignment with symbol H04M1, denoting "Substation equipment," suggests HTC's emerging emphasis on mobile phone software.

In the year 2016, many co-assigned pairs from the previous years disappear, and there is significant growth for patents co-assigned with G06F3 and G06F1. Especially, H04L27 and H04L43, which have the highest co-assignment frequency in the previous years, can no longer be found in Figs. 6 and 9. This seems to indicate that HTC's complete position shifting from a communication-oriented company to a software-oriented company.

This shift of R&D focus and position is more vividly illustrated in Fig. 10, where the co-assignment frequencies of symbol pairs are ranked for each year and the rank variations of the top ranking symbol pairs over the years are depicted. One may see that the co-assignment of G06F1 and H04M1 is rare before 2013 but the pair dramatically becomes the more prominent one after 2014. In contrast, those top ranking pairs before 2015 quickly sink into oblivion in 2016.

IV. DISCUSSION

One may argue that, if most patents have just a single classification symbol, the two-dimensional profile matrix does not provide additional information over the traditional one-dimensional bar chart. This is a valid argument but, fortunately, most patents do have plural classification symbols. The authors once analyzed 234,966 U.S. utility patents issued in 2012 and found that on the average each patent has 3.9 USPC symbols. This phenomenon is applicable to CPC symbols as well.

Another argument is that a patent's technology content may involve some portion that leads to the co-assignment of, not two, but three or more classification symbols. In other words, if a patent has classification symbols {C*i*, C*j*, C*k*}, this argument suggests that some portion of the patent's technology content is hidden not in the co-assignment of {C*i*, C*j*}, {C*j*, C*k*}, or {C*i*, C*k*}, but in the co-assignment of {C*i*, C*j*, C*k*}. This is a possible scenario but a high co-assignment frequency of {C*i*, C*j*, C*k*} should also be reflected in comparably high co-assignment frequencies of {C*i*, C*j*}, {C*j*, C*k*}, and {C*i*, C*k*}. This scenario, however, indeed suggests that some pairs of classification symbols are actually "noises." This issue should be investigated in the future to differentiate true signals from noises.

One may also argue that if a symbol has a high individual assignment frequency, it should also have high co-assignment frequencies with other symbols. In other words, this argument suggests that the traditional one-dimensional technology profile is good enough and there is no need to build a two-dimensional one. This argument, however, has no merit. For example, in Fig. 4, the symbol H04M1 has a rather high individual assignment frequency 216 but it is rarely co-assigned with other symbols. In addition, it is common that a symbol has higher co-assignment frequencies than its individual assignment frequency, such as the symbols H04N21, H04L5, H04L1, H04L12, H04H20 shown in Fig. 4.

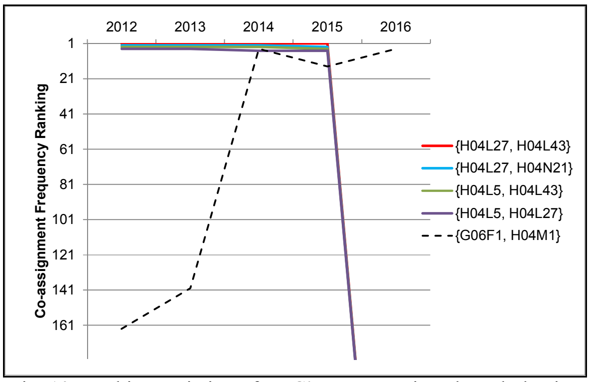

Fig. 10. Ranking variation of HTC's top co-assigned symbol pairs.

MDS is by no means the only way to process the profile matrix. Alternatively, as the profile matrix may be considered as an adjacency matrix, a network analysis tool such as VOSviewer [14] or Pajek [15] may be applied. Fig. 11 is the graphical representation of the 2016 profile matrix of Fig. 6 by VOSviewer, where thicker links indicate greater co-assignment frequencies, and larger nodes reflect symbols' greater overall frequencies (including both individually assigned frequencies and co-assigned frequencies).

V. CONCLUSION

This study notices that a patent's individual classification symbols only capture portions of the patent's technology content, and there may be some content "hidden" in the co-assignment of two or more symbols. This study therefore adopts a two-dimensional profile matrix for more comprehensively capturing the technology content of a patent portfolio. The element Mij of the profile matrix M counts the co-assignment frequency for symbols Ci and Cj, and the element Mii is the individual assignment frequency of symbol Ci, to the patents of the portfolio. The profile matrix therefore not only is an extension of the traditional one-dimensional patent classification analysis, but also provides a more complete picture of the portfolio's technology content.

Considering that the profile matrix captures more detailed information about a portfolio's technology content, the study conducts empirical observation to the company HTC's patent portfolio from 2012 to 2016. Based on the observation result, the profile matrix indeed may be utilized to reveal whether and how an entity shifts its technology development direction over time. HTC, however, is a medium sized company with a limited number of patents. It would be interesting to apply profile matrix to more sizeable entities such as an entire technology field to verify the effectiveness of profile matrix. Such endeavor is currently under way by the authors.

In addition to monitoring an entity's evolving technology development, for example, the profile matrices of different entities may be compared to detect the degree of similarity or relatedness between their portfolios. Then, companies having comparable or incomparable profile matrices may be potential targets for collaboration or mergers and acquisitions (M&A).

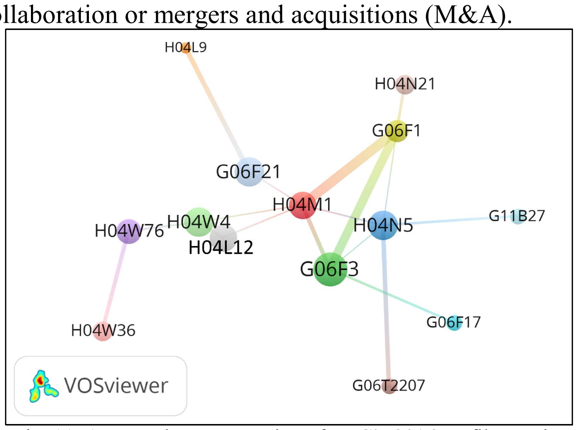

Fig. 11. A network representation of HTC's 2016 profile matrix.

Furthermore, for a technology field, multiple profile matrices for the field's portfolio at different times may be constructed and compared to identify emerging new technologies.

The rationale behind this application is that appropriate classification symbols for emerging technologies, especially those cross-disciplinary ones, may not be available from the classification scheme (until it is revised and new version is published in the future). Then, for these new technologies, the co-assignment of multiple symbols would be inevitable. Then, the profile matrix would be ideal to detect this kind of new technologies.

The detection of emerging technologies may be conducted as follows. Firstly, a series of profile matrices are constructed at successive instances of time for a technology field. Then, these profile matrices are compared to see if there is some element Mij has a significantly high growth rate. Please note that the co-assignment frequency growth rate should be significantly higher than the patents' growth rate so that the increase of the co-assignment frequency of symbols Ci and Cj is not an automatic outcome of the growing number of patents. After detecting this phenomenon, the patents having the co-assigned symbols Ci and Cj are reviewed to see what new technology is involved.

Whether this methodology really works is an interesting topic and, if it does work, the profile matrix would be a valuable tool in detecting emerging technologies. The authors are currently applying profile matrix in this manner to verify its validity in this respect.

REFERENCES

[1] World Intellectual Property Organization (WIPO), *Guide to the IPC*, 2012. Retrieved from http://www.wipo.int/classifications/ipc/en/guide/guide_ipc.pdf.

[2] European Patent Office and U.S. Patent and Trademark Office, *Guide to the CPC*, 2017. Retrieved from http://www.cooperativepatentclassification.org/publications/GuideToTheCPC.pdf.

[3] Y.-C. Chiu, H.-C. Lai, Y.-C. Liaw, and T.-Y. Lee, "Technological scope: diversified or specialized," *Scientometrics*, vol. 82, no. 1, pp. 37–58, Oct. 2009.

[4] J. H. Chen, S.-L. Jang, and S. H. Wen, "Measuring technological diversification: identifying the effects of patent scale and patent scope," *Scientometrics*, vol. 84, no. 1, pp. 265–275, 2009.

[5] The Organisation for Economic Co-operation and Development (OECD), *The measurement of scientific and technological activities using patent data as science and technology indicators: Patent manual*, 1994. Retrieved from http://www.oecd.org/sti/inno/2095942.pdf.

[6] L. Leydesdorff, "Patent classifications as indicators of intellectual organization," *Journal of the American Society for Information Science and Technology*, vol. 59, no. 10, pp. 1582–1597, 2008.

[7] U. Schmoch, F. Laville, M. Pianta, and G. Sirilli, "The measurement of scientific and technological activities:

using patent data as science and technology indicators," Patent Manual, 1994.

[8] A. B. Jaffe, "Technological Opportunity and Spillovers of R&D: Evidence from Firms Patents, Profits and Market Value," *The American Economic Review*, vol. 76, no. 5, pp. 984–1001, 1986.

[9] A. B. Jaffe, "Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers," *Res. Policy*, vol. 18, no. 2, pp. 87–97, 1989.

[10] M. Trajtenberg, R. Henderson, and A. Jaffe, "University versus corporate patents: A window on the basicness of invention," *Economics of Innovation and New Technology*, vol. 5, no. 1, pp. 19–50, 1997.

[11] R. C. McNamee, "Can't see the forest for the leaves: similarity and distance measures for hierarchical taxonomies with a patent classification example." *Res. Policy*, vol. 42, pp. 855–873, 2013.

[12] S. Breschi, F. Lissoni, and F. Malerba, "Knowledge-relatedness in firm technological diversification," *Res. Policy*, vol. 32, pp. 69–87, 2003.

[13] K. Wittenburg, and G. Pekhteryev, "Multi-dimensional comparative visualization for patent landscaping," in *Proc. Of BusinessVis Workshop*, 2015.

[14] N. J. V. Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2009.

[15] W. de Nooy, A. Mrvar, and V. Batagelj, *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press, 2018.