Exploring potential patent portfolios:

an integrated approach based on topic identification and correlation analysis

Xian Zhang, Haiyun Xu, Shu Fang, Zhengyin Hu

(Chengdu Library of the Chinese Academy of Sciences)

Leiden, The Netherlands
Sept .2, 2014

Outline

- 1. Background
- 2. Methodology
- 3. Empirical Study
- 4. Discussion & Conclusion

1. Background

- **About Patent portfolio**
  - Presented by Ernst (1998), to analyze the technological integration capabilities of enterprises, and the cross-field potential R&D collaborators
  - Whole is greater than the sum of its parts (Parchomovsky & Wagner, 2005)
  - Conceptually, there are three types of technology portfolio (Shin, 2013) , a relatively static, a dynamic contingent, a shift to new innovation
  - focused on the value assessments, applications, the volume and formation of patent portfolios within a given assignee, and etc.(Yue, 2013), but relatively little research on the potential patent portfolios identifying cross the assignees.

1. Background (continued)

- **Main existing approaches to patent portfolio**
  - The majority was undertaken on the basis of citation relationships (Hall,2000; Kuan,2013)
  - A Patent Priority Approach was raised by Su (Su,2009;2011) and improved by the followers (Pan, 2013)
  - Indicator-integrated systems, developing several indicators for drawing patent portfolio diagrams and performing portfolio analysis (Ernst, 1998; Wang,2009; Wu,2014)
  - Modification and application of Lanjouw-Schankerman model (Hu, 2007; Cao, 2012), and Black-Litterman model (Shin, 2013)

1. Background (continued)

- **Limitations**
  - focusing on the external characteristics of patent documents instead of the technical solutions
  - Still ambiguous
    - Citation relationship is manipulated by the others. The technology development context of the focal firm cannot be observed (Su, 2011).
    - Citation and co-citation is strongly criticized as they are not related to the context of the document (Kraslawski, 2006; Alcacer,2006).
    - In addition to citations and priority and even the classifications, the textual units are the better proxy of the technologic meanings of patent solutions (Leydesdorff, 2008).
    - Patent priority network is useful for the portfolios identifying within a given enterprise, but not helpful for that between different enterprises
    - There needs more robust models and more case studies conducted in different sectors, for the industry dependence of patents.

2. Methodology

- **Our objective**
  - develop a framework to identify the potential patent portfolios
  - so as to identify potential patent pools
  - and find the clues for the further cross-assignee collaborators

- **Framework of our solution**
  - Technical topics extraction
    - IPC, DMC, technological terms
  - Correlation analysis
    - Constructing association networks
    - Clustering and visualizing
  - Interpreting the maps

2. Methodology (continued)

- **Main steps**
  - (1) Extracting the technology topics of the patent solutions
    - ✓ In two dimensions
      - Structural information: IPC, DMC, reflecting the integrity of the technical structure
      - Non-structural information: technology terms (NLP by TDA)
  - (2) Correlation analysis
    - ✓ Constructing association networks
      - Type 1: based on the assignees and technology topics, to investigate the correlations of these topics; identifying the potential technology portfolios.
      - Type 2: cross association network based on the topics and assignees, to investigate the potential partners tied by the common patenting theme
    - ✓ Clustering in topics, Computing the similarities and visualizing with NetDraw.
  - (3) Interpreting the maps
    - ✓ Combining the above quantitative analysis and the qualitative opinion from technology experts. identifying the potential patent portfolio and the potential cooperation opportunities for the patent assignees.

3. Empirical Study

- **Data source & tools**
  - Derwent Innovation Index (retrieval date: 2013-11-20)
  - Thomson Data Analyzer
  - NetDraw

- **Sample records**
  - Chemistry patents filed by the Chinese Academy of the Sciences (CAS) during 1992-2012
  - Defining the field of the chemistry: according to WIPO IPC-technology concordance table
  - 27401 records hit

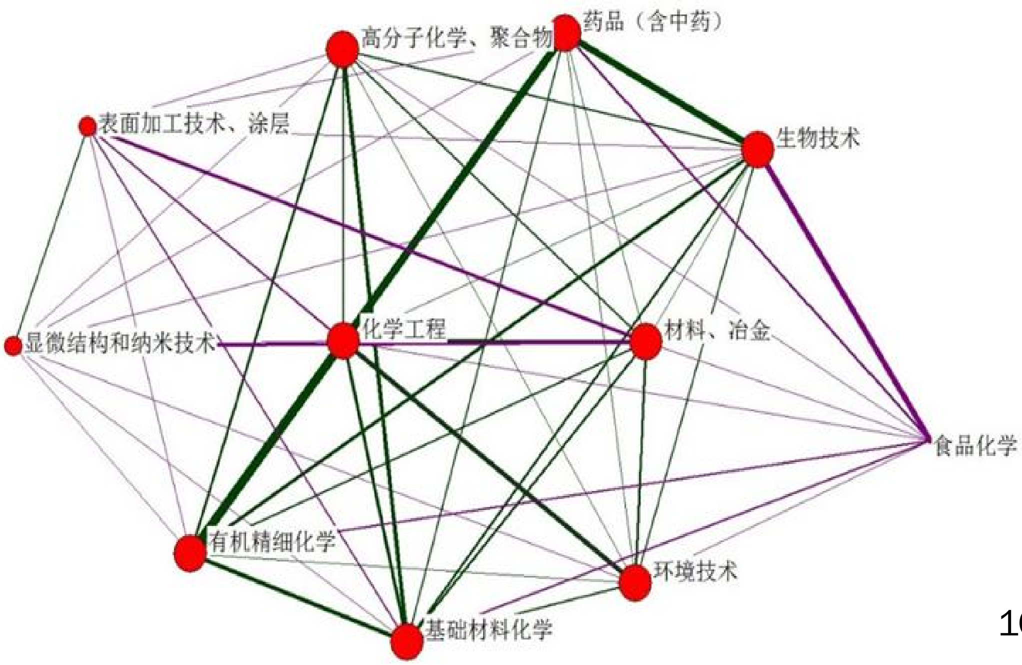
Chemistry fields in WIPO IPC-technology concordance table (revised by Jan. 2013)

Area	field	IPC Code
Chemistry	Organic fine chemistry	A61K-008，A61Q，C07B，C07C，C07D，C07F，C07H，C07J，C40B
	Biotechnology	C07G，C07K，C12M，C12N，C12P，C12Q，C12R，C12S
	Pharmaceuticals	A61K-006，A61K-009，A61K-031，A61K-033，A61K-035，A61K-036，A61K-038，A61K-039，A61K-041，A61K-045，A61K-047，A61K-048，A61K-049，A61K-050，A61K-051，A61K-101，A61K-103，A61K-125，A61K-127，A61K-129，A61K-131，A61K-133，A61K-135，A61P
	Macromolecular chemistry, polymers	C08B，C08C，C08F，C08G，C08H，C08K，C08L
	Food chemistry	A01H，A21D，A23B，A23C，A23D，A23F，A23G，A23J，A23K，A23L，C12C，C12F，C12G，C12H，C12J，C13B-010，C13B-020，C13B-030，C13B-035，C13B-040，C13B-050，C13B-099，C13D，C13F，C13J，C13K
	Basic materials chemistry	A01N，A01P，C05B，C05C，C05D，C05F，C05G，C06B，C06C，C06D，C06F，C09B，C09C，C09D，C09F，C09G，C09H，C09J，C09K，C10B，C10C，C10F，C10G，C10H，C10J，C10K，C10L，C10M，C10N，C11B，C11C，C11D，C99Z
	Materials, metallurgy	B22C，B22D，B22F，C01B，C01C，C01D，C01F，C01G，C03C，C04B，C21B，C21C，C21D，C22B，C22C，C22F
	Surface technology, coating	B05C，B05D，B32B，C23C，C23D，C23F，C23G，C25B，C25C，C25D，C25F，C30B
	Micro-structure and nano-technology	B81B，B81C，B82B，B82Y
	Chemical engineering	B01B，B01D-001，B01D-003，B01D-005，B01D-007，B01D-008，B01D-009，B01D-011，B01D-012，B01D-015，B01D-017，B01D-019，B01D-021，B01D-024，B01D-025，B01D-027，B01D-029，B01D-033，B01D-035，B01D-036，B01D-037，B01D-039，B01D-041，B01D-043，B01D-057，B01D-059，B01D-061，B01D-063，B01D-065，B01D-067，B01D-069，B01D-071，B01F，B01J，B01L，B02C，B03B，B03C，B03D，B04B，B04C，B05B，B06B，B07B，B07C，B08B，C14C，D06B，D06C，D06L，F25J，F26B，H05H
	Environmental technology	A62C，B01D-045，B01D-046，B01D-047，B01D-049，B01D-050，B01D-051，B01D-052，B01D-053，B09B，B09C，B65F，C02F，E01F-008，F01N，F23G，F23J，G01T

3. Empirical Study (continued)

- **(1) Exploring the linkages between the main fields of Chemistry area**
  - E-I (external-internal) Index is used
    - to test the agglomeration or dispersion of the networks.
    - takes the number of ties of group members to outsiders, subtracts the number of ties to other group members, and divides by the total number of ties.
    - E-I Index value $\in [-1, +1]$

- **Ties in purple: inter-field relationships**
- **Ties in green: inner-field relationships**
- **E-I index = 0.094**



3. Empirical Study (continued)

- **(1) Exploring the linkages between the main fields of Chemistry area (continued)**
  - Derwent Manual Code (DMC) is employed as the proxy of topics of patent solution.
  - Girvan–Newman algorithm is used
    - to detect communities by progressively removing edges from the original network, to test the agglomeration or dispersion of the networks.
    - It is a kind of exploration cluster method in complex systems, focusing on edges that are most likely "between" communities.
    - The connected components of the remaining network are the communities.
    - Main steps for community detection:
      - The betweenness of all existing edges in the network is calculated first.
      - The edge with the highest betweenness is removed.
      - The betweenness of all edges affected by the removal is recalculated.
      - Steps 2 and 3 are repeated until no edges remain.

Finding 1:

- The linkages between the chemical engineering, medicine (including traditional Chinese Medicine) and organic fine chemicals are closer and stronger, while the food chemistry, microstructure and nanotechnology, and surface processing and coating are linked weakly.
- The CAS's patents on CE can be clustered into 6 categories: (i) polymer, plastic; (ii) general chemistry; (iii) catalyst; (iv) drugs; (v) agriculture; (vi) semiconductor, circuit, fireproofing, ceramics, cement, electrochemical.
- Polymers, plastics is located in the center of the network, including: monomer, concentration, polymerization, natural polymer, addition polymer, condensation polymer, inorganic polymer, polymer blending, water dispersion system, additive, property, analysis, testing, control, polymerization, polymer modification, polymer processing, application of polymers.
- Core node of the network: the application of polymers

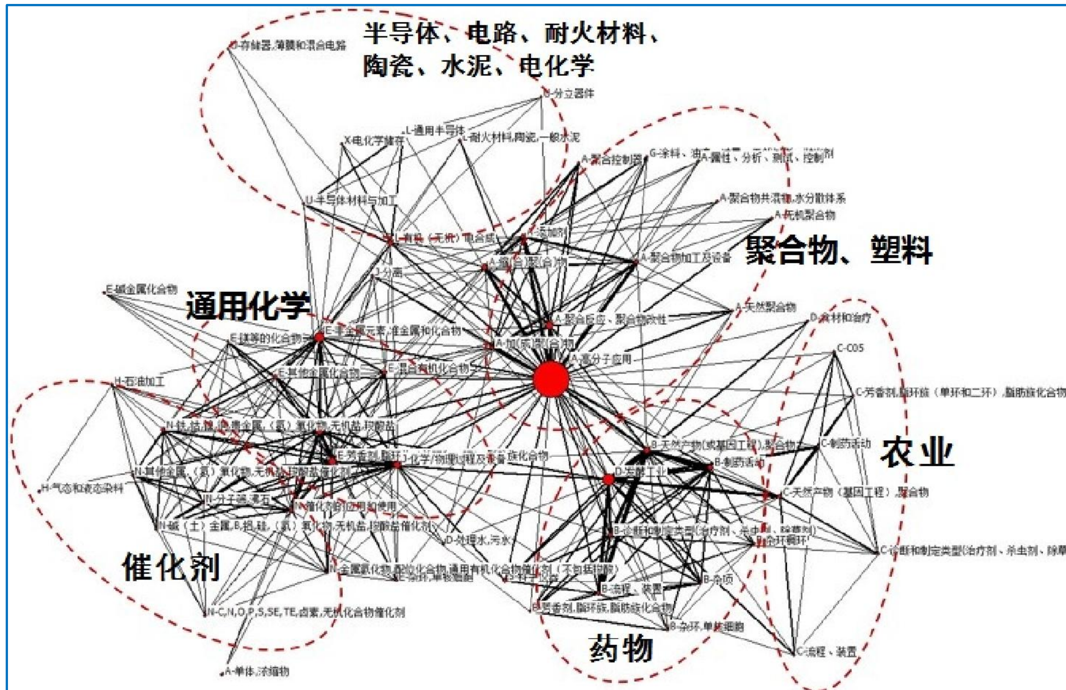


Figure 1: Topic co-occurrence network of chemistry patent of CAS (based on DMC, frequency≥100)

3. Empirical Study (continued)

- **(2) Exploring the opportunities or patent portfolio**
  - Focusing on the field of application of polymers (core node of Figure 1): 409 hits
  - Technological term extraction: from the textual elements (title, abstract, claims), NLP (natural language phrase) by TDA
  - Association network: Co-occurrence of technological terms
  - Size of the nodes: the frequency of occurrence of the topics
  - The thickness of the links: the strengthen of the relationship

Finding 2:

- CAS's polymer-application patents can be roughly summed up into 6 areas: biological polymers, synthetic resin, conductive polymers, engineering plastics, fertilizers and polyamide system.
- Together with the technical experts, we investigate the possible solutions to patent portfolios.
- For example, one perspective is considering the possible applications of one technology in different products or processes. Another perspective is considering the possible combination of different techniques around a product or a process.

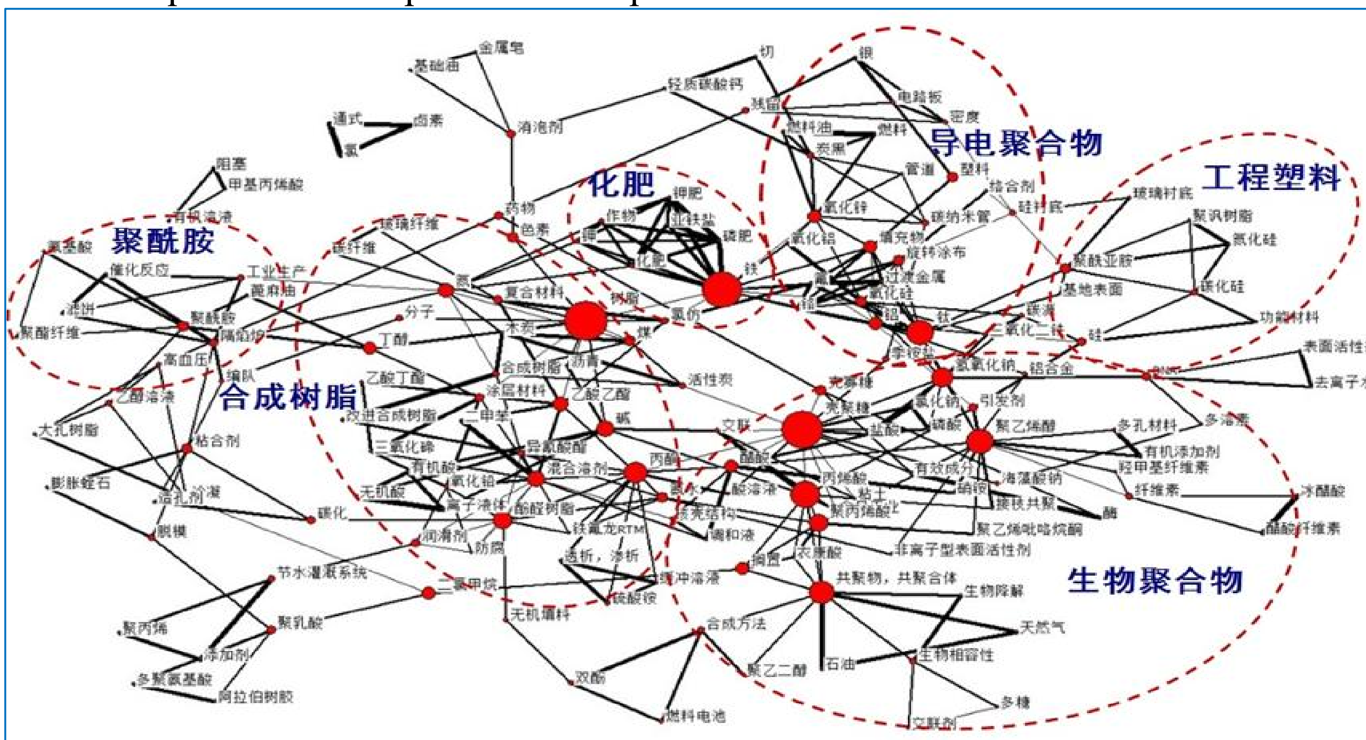


Figure 2: Potential patent portfolios of polymer applications of CAS

3. Empirical Study (continued)

- **(3) Exploring the potential cooperators**
  - Selecting the Top 10 assignees on polymer applications
  - Selecting the key topics
  - Performing a cross co-occurrence network
  - Together with the technical experts, finding the presence of multiple technique-cross area between or among these assignees.
  - These common technical means there may be highly potential cooperation for these institutes.

Finding 3:

- Together with the technical experts, on the basis of the common technical means among these institutes, we have discussed some of the most promising cooperation in detail. For example:
  - ✓ A, B, D, F: Synthesis and application of phenolic resin and other resin
  - ✓ A, B, C, G, F: preparation and application of polyethylene glycol
  - ✓ B, D, E: preparation and application of chitosan
  - ✓ A, E: Aluminum metal material and its preparation
  - ✓ B, C: Preparation and application of coating material
  - ✓ A, C, F: Preparation and application of polyvinyl alcohol

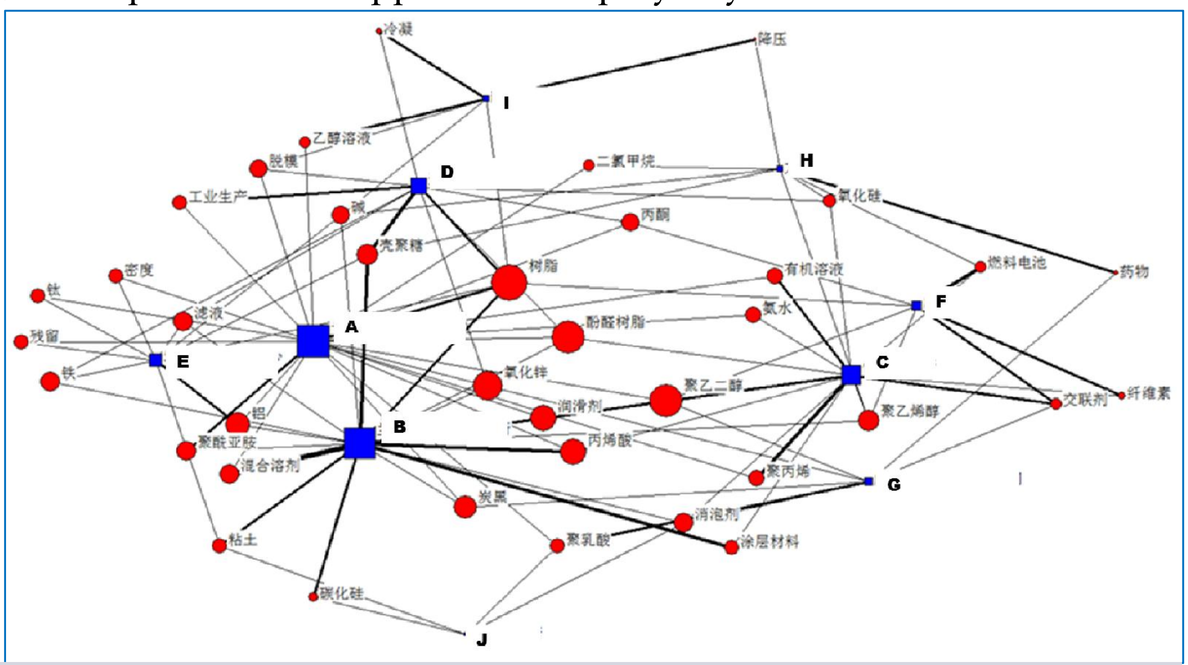


Figure 3: Potential partners in polymer application of CAS

4. Discussion & Conclusion

- **Achievements**
  - Also can be applied by enterprise and other research institutes like universities.
  - The findings may not only be valuable to scholars but also to policy makers and practitioners.
  - Enabling the industries and well-connected institutes to develop higher impact patent portfolios.

- **Improvement on the road ahead**
  - More assessments of experts should be integrated.
  - In addition to topic terms and DMCs, more textual elements and semantic analysis should be considered in this approach.
  - More case studies conducted in different technical fields are needed.

Part of References

- Alcacer J, Gittelman M. (2006). Patent citations as a measure of knowledge flows: the influence of examiner citations. The Review of Economics and Statistics, 88,4, 774-779.
- Cao C., Hu YJ. (2012). Patent Portfolio Evaluation: an Empirical Study Based on Drug Patent. Science and Technology Management Research, 13, 174-177.
- ERNST H. (1998). Patent portfolios for strategic R&D planning. Journal of Engineering and Technology Management, 4, 279-308.
- Girvan M, Newman M E J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99,12: 7821-7826.
- Hall Bronwyn H. (2001). Market Value and Patent Citations: A First Look[R].
- Kraslawski A. (2006). Semantic Analysis for Identification of Portfolio of R&D projects - Example of Microencapsulation. 16th European Symposium On Computer Aided Process Engineering And 9th International Symposium On Process Systems Engineering, 21, 1905-1910.
- Kuan, CH., Huang, MH, Chen, DZ. (2013). Capturing and Tracking Performance of Patent Portfolio Using h-Complement Area Centroid. IEEE Transactions on Engineering Management, 60,3,496-505.
- Leydesdorff L. (2008). Patent Classifications as Indicators of Intellectual Organization. Journal of the American Society for Information Science and Technology, 59, 10,1582–1597.

Part of References (continued)

- Newman M E J., Girvan M. (2004). Finding and evaluating community structure in networks. Physical review E, 69,2, 026113.
- Pan Y., Lu ZP. (2012). Patent Priority Network: A New Patent Portfolio Analysis Method. Library and Information Service, 56, 16, 97-101.
- Parchomovsky,G., Wagner P. (2005). Patent Portfolios[R]. University of Pennsylvania Law Review.
- Shin, J; Coh, BY; Lee, C. (2013). Robust future-oriented technology portfolios: Black-Litterman approach. R & D MANAGEMENT,43,5,409-419.
- Su, F.P., Lai, K.K., Sharma, R.R.K. and Kuo, T.H. (2009). Patent priority network: linkage patent portfolio strategic goals. J. Am. Soc. Inf. Sci. Technol., 60, 2353-2361.
- Su, FP; Shih, HC. (2011). Building Core Technological Competence: Patent Portfolio Perspective. 2011 Proceedings Of Picmet 11: Technology Management In The Energy-Smart World (Picmet), Jul 31-Aug 4.
- Su, F. P., Yang, W.G. and Lai, K.K. (2011). A Heuristic Procedure to Identify the Most Valuable Chain of Patent Priority Network. Technol. Forecast. Soc. Chang., 78,319-331.
- WIPO. (2013). IPC - Technology Concordance Table. http://www.wipo.int/ipstats/en/statistics/technology_concordance.html.[EBOL], 2013-1.
- Yue, XP. (2012). Inter-firm Patent Portfolio Scale Strategy: Based on Cournot Mode. Journal of Intelligence, 31, 11,118-122,135.

Acknowledgement

- This work is supported by the Intellectual Property Information Service Program and the West Light Foundation Program of the Chinese Academy of Sciences, and the National Social Science Foundation of China.

Thanks for your attention!

Xian Zhang
(Chengdu Library of Chinese Academy of Sciences)

Add: P.O. Box 20, No. 16 South Sec. 2 Yihuan Rd., Chengdu 610041, P.R. China
Phone: +86-28-85228846
Fax: +86-28-85220439
Email: zhangx@clas.ac.cn