A Two-Layer Deep Learning Approach for R&D Partner Recommendation in the Self-Driving Vehicle Industry

Juite Wang and Ying-Pei Kao
Graduate Institute of Technology Management, National Chung Hsing University, 145 Xingda Rd, Taichung, Taiwan, Republic of China

Keywords: Self-Driving Vehicles, Partner Recommendation, Patent Analytics, Social Network Analysis, Deep Neural Networks, Graph Neural Networks.

Abstract: This study presents a two-stage deep learning framework for recommending strategic R&D partners in the self-driving vehicle (SDV) industry. Leveraging 165,775 U.S. patent applications from 2015 to 2023, we constructed a co-patent network and extracted node, edge, and topological features to represent organizational attributes, collaboration intensity, and network structure. These features were integrated using a hybrid Graph Neural Network (GNN) and Deep Neural Network (DNN) architecture to predict future collaborations. The model achieved high predictive performance (accuracy = 96.65%, precision = 70.83%, recall = 66.92%, F1 = 68.82%, and AUPRC = 78.93%) and demonstrated its ability to identify both established and emerging partners. Community detection revealed influential clusters anchored by firms like Toyota and Hyundai. Case analyses showed that the model can recommend both historical and emerging R&D partners. Compared to prior work, this study contributes a scalable, data-driven approach that incorporates deep structural and semantic signals to improve partner selection accuracy. The framework advances patent analytics by linking network-based learning with partner recommendations, offering practical implications for R&D planning in complex technology-based industries.

1 INTRODUCTION

The automotive industry is undergoing a profound transformation driven by self-driving vehicle (SDV) technologies. These advancements build on artificial intelligence (AI), big data, 5G, Internet of Things (IoT), and computer vision. Leading automakers and technology firms alike are investing in SDV development, while also seeking external partners to accelerate innovation. Strategic R&D collaboration has thus become a critical issue in the competitive landscape.

The complexity of SDV innovation requires coordination across hardware, software, and infrastructure layers. Companies must integrate capabilities in sensing technologies, real-time decision-making algorithms, high-definition mapping, and connectivity. This interdisciplinary nature not only amplifies technical challenges but also elevates the strategic importance of effective partner selection. As competition intensifies, firms that can identify and secure partnerships with technically aligned collaborators will gain substantial advantages in innovation speed and technological scope.

Despite this urgency, systematic methods for identifying suitable R&D partners remain limited. Traditional approaches often rely on prior relationships, managerial intuition, or firm-level reputational signals. While helpful, these heuristics overlook valuable insights embedded in structured data sources such as patent filings. Patent data reveal not only technological strengths but also patterns of past cooperation, citation influence, and topical complementarities. By transforming patent co-application information into a relational network, one can better understand the evolving landscape of collaboration and anticipate future alliance opportunities.

SDV development requires integration of diverse technical domains. No single firm can master all areas, making partner selection crucial for acquiring complementary capabilities and accelerating time-to-market. While prior collaborations provide valuable signals, identifying future partners based on technical fit and network dynamics remains a challenging

Wang, J. and Kao, Y.-P.
A Two-Layer Deep Learning Approach for R&D Partner Recommendation in the Self-Driving Vehicle Industry.
DOI: 10.5220/0013683100004000

In *Proceedings of the 17th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2025) - Volume 1: KDIR*, pages 60-66

problem. This challenge is particularly salient in the SDV context, where the rapid pace of change demands a forward-looking approach to alliance formation.

Patent data offer a valuable lens for analyzing R&D collaboration. They reflect firms' technological focus and inventive activity and, when co-assigned, signal formalized joint efforts. Studies on R&D partner selection using patent data can be broadly categorized into three methodological approaches: bibliometric analysis (Geum et al., 2013; Wang, 2012), link prediction (Yan & Guns, 2014), and textual analysis (Kim et al., 2020; Wang et al., 2017). Yet, significant research gaps remain. First, most prior studies on partner selection have relied on static indicators or network metrics. These methods fail to capture the complex, non-linear interactions among structural features and R&D collaboration potential. Second, few studies integrate deep learning with patent analytics, despite growing evidence that neural networks can improve pattern recognition in structured and semi-structured data.

This study aims to develop a two-stage deep learning model to predict potential R&D partnerships in the self-driving vehicle industry. Patent data from the United States Patent and Trademark Office (USPTO) were collected. Co-applicant patents were used to build an R&D collaboration network. Bibliometric indicators, link prediction measures, and social network analysis were used to generate features. A two-stage deep learning model, comprising a GNN and a DNN, was trained to predict collaboration links.

According to the obtained results, the model achieved acceptable predictive performance (F1 score = 68.82, AUPRC = 78.93) by leveraging node, edge, and topological features extracted from a co-patent network. Network analysis revealed influential clusters centered on firms like Toyota and Hyundai, and the partner recommendation system successfully identified both established and emerging collaborators, offering practical insights for strategic alliance planning.

This study contributes to the literature by (1) introducing a deep learning-based framework that unifies node, edge, and topological features; (2) applying GNN and DNN models to large-scale patent co-application data; and (3) demonstrating how the system can generate interpretable partner recommendations grounded in empirical collaboration signals.

Section 2 reviews the literature on SDVs and partner selection. Section 3 outlines the data and methodological framework. Section 4 presents the empirical results. Section 5 concludes with future directions.

2 LITERATURE REVIEW

2.1 Self-Driving Vehicles

SDVs, also referred to as autonomous vehicles, represent a convergence of several advanced technologies, including machine learning, sensor fusion, computer vision, and real-time decision-making algorithms (Gwak et al., 2019). Research on SDVs has examined their potential to reduce accidents, improve traffic flow, and enhance mobility for underserved populations (Ryan, 2020). Technological classifications based on SAE levels have been widely adopted, and attention has increasingly turned toward regulatory, infrastructural, and ethical considerations.

In addition to engineering and legal aspects, researchers have explored the business and ecosystem implications of SDVs. These include the emergence of mobility-as-a-service (MaaS), the integration of SDVs into smart cities, and the disruption of incumbent value chains (Vosooghi et al., 2019). However, many of these studies treat technological development as a monolithic process, overlooking the collaborative innovation dynamics that drive SDV advancements.

2.2 Partner Selection and Patent Analytics

In innovation studies, partner selection has long been recognized as a determinant of alliance success (Shah & Swaminathan, 2008). Firms enter R&D collaborations to access complementary capabilities, share costs, and reduce uncertainties in technological exploration (Makri et al., 2010). Effective partner selection thus requires evaluating not only organizational fit but also technical and relational compatibility.

Studies on R&D partner selection using patent data can be broadly categorized into three methodological approaches: bibliometric analysis, link prediction, and textual analysis. Bibliometric methods use structured metadata such as citations, classifications, and assignee information to assess firms' technological strength and collaborative potential (Geum et al., 2013; Wang, 2012), though they are limited by citation delays and outdated classification systems (Kim et al., 2020). Link prediction approaches (Yan & Guns, 2014) leverage co-patenting or co-authorship networks to forecast partnerships based on network topology, but they often overlook non-structural factors that influence collaboration decisions (Song et al., 2016). Textual analysis applies NLP techniques to extract semantic insights from patent texts, using methods like Subject-Action-Object (SAO) structures or embedding models to identify technological similarity (Kim et al., 2020; Wang et al., 2017), though earlier approaches lacked deep contextual understanding.

To address the limitations of single-method approaches, a growing number of studies have adopted hybrid models that combine multiple analytical perspectives. For instance, Yoon and Song (2014) integrated text mining and morphological analysis with bibliometric indicators to map technological opportunities and assess partner suitability. Park et al. (2015) employed latent semantic analysis (LSA) in conjunction with bibliographic coupling to evaluate technological proximity among firms. Song et al. (2016) extracted action-object structures from patent texts to assess technological needs and then applied three patent-based indicators to evaluate candidate partners. Jee and Sohn (2020) used latent Dirichlet allocation (LDA) topic modeling to identify firms working in complementary areas, followed by bibliometric analysis to evaluate technological strength and relatedness. Qi et al. (2022) applied LDA to both scientific publications and patent documents to extract topic-level similarities and then employed link prediction methods to recommend partners based on innovation capacity and openness. More recently, Wang et al. (2025) constructed three interconnected network layers and introduced three metrics, technological similarity (patent citation metric), technological heterogeneity (patent indicators), and collaboration status (based on link prediction indicators), to recommend potential R&D partners.

Although these hybrid approaches represent a meaningful advance over single-method models, important limitations remain. Most studies do not fully integrate all three core dimensions, textual analytics, bibliometric analysis, and network-based link prediction, within a unified analytical framework. Moreover, few adopt advanced machine learning methods capable of modeling complex, non-linear relationships among diverse features. This methodological fragmentation restricts the capacity of existing models to reflect the multifaceted and dynamic nature of R&D collaboration.

3 DATA AND METHODOLOGY

3.1 Data Collection and Preprocessing

This study collected 197,203 patent applications related to self-driving vehicles from the United States Patent and Trademark Office (USPTO) database, spanning from 2001 to October 2023. To align with the goal of recommending organizational R&D partners, all patents filed by individual applicants were excluded. The final dataset consisted of 165,775 organizational patents. As shown in Figure 1, the number of SDV-related patent applications has steadily increased each year, suggesting a growing market and technological momentum in this domain. The decline in 2023 filings reflects a lag due to the required publication and examination period for patent applications.

Figure 2 further shows that co-patent applications, those involving collaborations between organizations, have also grown since 2013. This trend highlights a rising demand for inter-organizational cooperation in the SDV industry. Notably, 2015 marks a critical milestone, with Tesla launching its third-generation electric vehicle (Model X), signaling the commercial viability of autonomous driving technologies. Accordingly, this study focuses on the period from 2015 to 2023, splitting the data into two segments: Period 1 (2015–2020) and Period 2 (2021–2023), to support time-aware modeling for potential R&D partner recommendation.

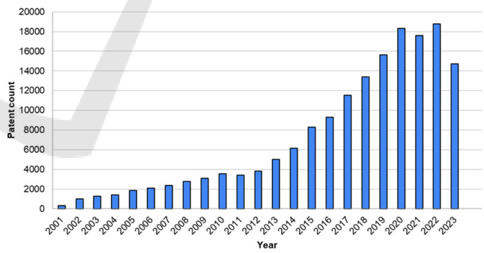

Figure 1: Yearly trend of patent applications.

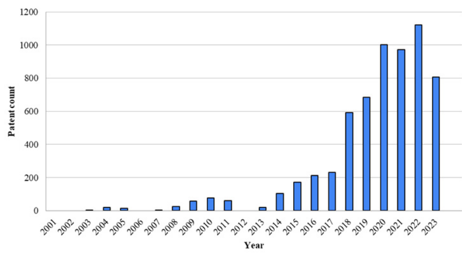

Figure 2: Yearly trend of collaborative patents.

3.2 Methodology

The methodological framework of this study, illustrated in Figure 3. From the applicant information in patent records, a collaboration network was constructed to represent co-application relationships between organizations. Various indicators were computed and categorized into three feature groups: node features, edge features, and topological features. These served as inputs for a deep learning model based on Park and Geum (2022), which integrates a Graph Neural Network (GNN) and a Deep Neural Network (DNN). In this framework, node features are processed by the GNN to generate structural embeddings, which are combined with edge and topological features in the DNN for link prediction. This two-stage approach supports accurate prediction and practical recommendation of R&D partners in the SDV domain.

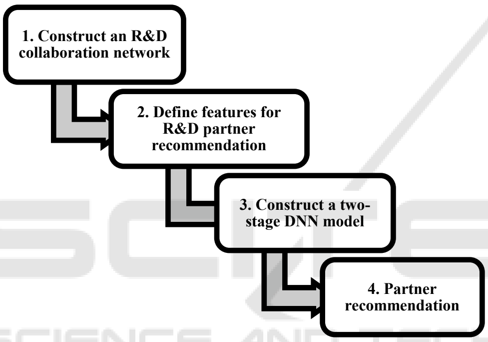

Figure 3: Research framework.

3.2.1 Constructing the Collaboration Network

To construct the collaboration network, each undirected edge represented at least one jointly filed patent between two organizations. The adjacency matrix was defined such that if firms and co-applied for at least one patent between 2015 and 2020, and otherwise, where edge weights reflected the number of co-patents. This network served as the foundation for feature generation and potential collaboration prediction tasks.

3.2.2 Defining Features for the DNN Model

Node features (organizational-level) are categorized into three categories.

Technological strength (Geum et al., 2013) includes:

- Patent quantity: number of application and granted patents;
- Patent quality: forward citations of application and granted patents;
- R&D resources: number of inventors and granted-to-application ratio;
- Technological presence: share of total patents and simple patent family ratio;
- Technological impact: proportion of highly cited (top 10%) patents.

R&D openness (Geum et al., 2013) captures:

- Organizational openness: share of co-invented patents;
- Inventor and assignee openness: average number of nationalities per patent.

Other variables developed in this study include:

- Organization nationality (via one-hot encoding);
- Technological fields: A vector indicating the number of patents held by the organization in each CPC subclass.

Edge features (dyadic-level) describe collaborative strength and synergy and consist of two categories.

R&D Linkage (Geum et al., 2013) includes:

- Joint R&D capability: Share of patents co-invented with partners out of the organization's total co-invented patents.
- Joint application intensity: Share of patents jointly filed with partners out of all co-invented patents across organizations.
- Joint invention intensity: Share of patents jointly invented with partners out of all co-invented patents across organizations.

Collaborative Synergy (Geum et al., 2013) includes:

- Knowledge inflow: Ratio of the average backward citations in partner co-invented patents to the organization's overall average.
- Knowledge criticality: Ratio of the average forward citations in partner co-invented patents to the organization's overall average.
- Knowledge similarity: Sum of shared citations between the organization and its partner, normalized by the number of application patents held by both.

Topological features (network-level) were based on several link prediction indicators (Mutlu et al., 2020) and network centrality measures (Hansen et al., 2011):

- Link similarity: common neighbors, shortest paths, Jaccard index, Salton index, Sorensen index, preferential attachment, resource allocation index, Adamic-Adar index, Katz index.
- Node centrality: degree, closeness, and betweenness centrality.

3.2.3 Building a Two-Stage DNN Model

The two-stage deep learning model (Figure 4) consisted of:

- Stage 1: GNN embedding: GraphSAGE (Hamilton et al., 2017) aggregated each node's features and neighborhood context into vector embeddings that preserved local topology and semantic content.
- Stage 2: DNN classifier: The embeddings from the GNN were concatenated with edge and topological features and input to a DNN. The DNN used two hidden layers with ReLU activation, dropout, and focal loss to address class imbalance. Training was conducted on the 2015–2020 data, and predictions were validated against 2021–2023 collaborations.

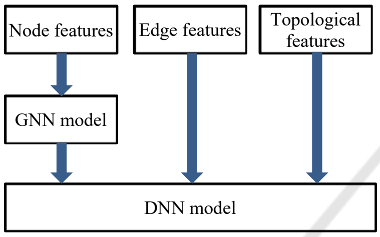

Figure 4: Two-stage DNN model.

Negative samples were created by randomly sampling unlinked firm pairs to balance the dataset for binary classification.

3.2.4 Partner Recommendation

The final output was an R&D partner recommendation list for each focal organization. Candidate firms were ranked based on their predicted link probabilities. The top-N ranked entities represented high-potential partners with strong structural and technical fit. These results offer practical guidance for alliance planning and partner identification in the self-driving vehicle innovation ecosystem.

4 RESULTS

4.1 Two-Stage Deep Learning Performance

The model was trained on 28,203 organization pairs constructed from 238 firms that collaborated in both periods. Data were split 70:30 into training and test sets. Using PyTorch, the two-stage deep learning framework was implemented: GNN embeddings were generated from node features, and then combined with edge and topological features to train the DNN. Hyperparameters were optimized using Optuna. The best configuration included a learning rate of 0.0089, three hidden layers in both GNN and DNN, Leaky ReLU for GNN activation, ReLU for DNN, and a MultiStepLR scheduler.

Model performance on the test set showed high accuracy (96.65%), precision (70.83%), recall (66.92%), and F1 score (68.82%). The area under the precision-recall curve (AUPRC) was 78.93, indicating strong performance across varying thresholds.

To assess the contribution of each feature set, separate models were trained using node, edge, or topological features alone. Results showed that while all single-feature models achieved reasonable accuracy, the integrated model outperformed them across all metrics. Notably, the GNN model yielded an F1 score of 58.87% and AUPRC of 72.70%, while edge and topological features were slightly lower. Combining all three feature sets enhanced precision and robustness.

4.2 Collaboration Network and Cluster Analysis

Community detection on the collaboration network revealed three dominant clusters:

- **Cluster 1:** Centered on Toyota, included 160 organizations, where Toyota, Honda, and Denso emerged as key nodes across centrality metrics.
- **Cluster 2:** Led by Hyundai, comprising 101 entities including Kia, LG, and Samsung, with strong interconnectivity among Korean firms.
- **Cluster 3:** Represented by Volkswagen, BMW, and Valeo, focusing on European and U.S. collaborations.

Centrality measures showed Toyota and Hyundai held the highest degree, betweenness, and eigenvector centralities within their clusters, confirming their influence.

4.3 Predicted R&D Collaboration Network and Partner Recommendation

Based on the trained two-stage model, this study predicted potential R&D collaborations for the 2024–2026 period using patent data from 2021–2023. A total of 590 likely collaboration pairs were identified, exceeding the 572 observed links in the prior network. Community detection and centrality analysis were applied to the predicted network, revealing new clusters and high-potential organizations.

- **Cluster 1:** Toyota remained central with the highest degree, betweenness, and eigenvector centralities, followed by Honda, Denso, Jtekt, and Stanford University.
- **Cluster 2:** Hyundai led the second-largest cluster, with Kia, Samsung, and Korea Advanced Institute of Science and Technology also prominently involved.
- **Cluster 3:** Emerging leaders included Increment P Corp and Pioneer Corp, along with research institutions such as Fraunhofer and Northeastern University.

Visualization of the predicted collaboration network (Figure 5) illustrated the structural role of core firms and highlighted new connections. Node size reflected the number of predicted links, while colors indicated cluster memberships.

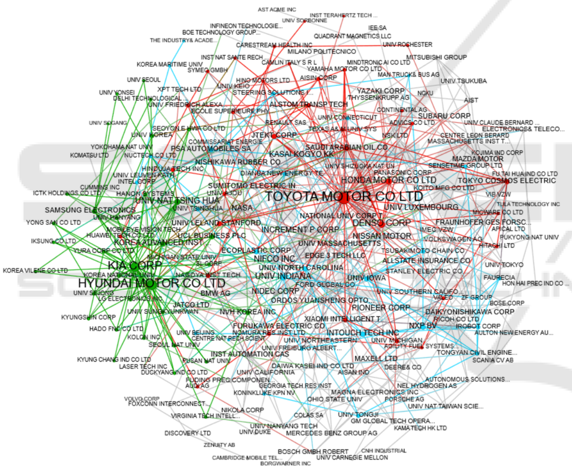

Figure 5: Predicted R&D collaboration network.

For example, the Toyota case revealed 35 predicted partners—most of which were previous collaborators such as Denso and Subaru, suggesting ongoing cooperation. However, new recommendations like Tokyo Cosmos Electric and Pioneer Corp were also identified based on CPC complementarity. Tokyo Cosmos Electric, with strengths in B60R and H05B, shares only one CPC class with Toyota but aligns with prior Toyota partners like Kojima Ind Corp, suggesting technical complementarity.

Recommended partners were categorized into:

- **Existing Partners:** reinforcing current alliances through shared CPC domains (e.g., Denso, Subaru).
- **New Partners:** identified based on complementary CPC profiles and inferred similarity to Toyota's prior partners (e.g., Tokyo Cosmos Electric, Univ Milano Politecnico).

These results demonstrate the model's capacity to uncover both strategic extensions of existing partnerships and novel opportunities based on technology fit.

5 CONCLUSIONS

This study developed a DNN framework to predict potential R&D partnerships in the self-driving vehicle industry based on patent data. Bibliometric indicators, link prediction measures, and social network analysis were used to generate features. A two-stage deep learning model, comprising a GNN and a DNN, was trained to predict collaboration links. The model achieved acceptable predictive performance by leveraging node, edge, and topological features extracted from a co-patent network. The obtained results revealed influential clusters centered on firms like Toyota and Hyundai, and the partner recommendation system successfully identified both established and emerging collaborators, offering practical insights for strategic alliance planning.

The limitations of this study are discussed as follows. First, this study analyzed only SDV related patent data in the USPTO database. Therefore, it is not cover patents related to SDV in other countries. For example, BYD Auto, one of the world's largest electric vehicle manufacturers in China, has files most of their patents in China. There are only 15 patents filed in US according to the collected patent data in this study. Second, this study is restricted to uncover R&D collaborations using patent data. It may not be able to find R&D collaboration in other forms, such as strategic alliance, new venture, and licensing (Schilling, 2012). Future research can include broader data sources that can yield a more comprehensive view of R&D collaboration in industries.

ACKNOWLEDGEMENTS

This research is supported by Grant No. NSTC 113-2221-E-005-076 -MY2 from the National Science and Technology Council, R.O.C.

REFERENCES

Geum, Y., Lee, S., Yoon, B., & Park, Y. (2013). Identifying and evaluating strategic partners for collaborative R&D: Index-based approach using patents and publications. *Technovation*, *33*(6-7), 211-224.

Gwak, J., Jung, J., Oh, R., Park, M., Rakhimov, M. A. K., & Ahn, J. (2019). A Review of Intelligent Self-Driving Vehicle Software Research [Review]. *Ksii Transactions on Internet and Information Systems*, *13*(11), 5299-5320. https://doi.org/10.3837/tiis.2019.11.002

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. arXiv:1706.02216. Retrieved June 01, 2017, from https://ui.adsabs.harvard.edu/abs/2017arXiv170602216H

Hansen, D. L., Shneiderman, B., & Smith, M. A. (2011). Chapter 5 - Calculating and Visualizing Network Metrics. In D. L. Hansen, B. Shneiderman, & M. A. Smith (Eds.), *Analyzing Social Media Networks with NodeXL* (pp. 69-78). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-382229-1.00005-9

Jee, S. J., & Sohn, S. Y. (2020). Patent-based framework for assisting entrepreneurial firms' R&D partner selection: Leveraging their limited resources and managing the tension between learning and protection. *Journal Of Engineering And Technology Management*, *57*, 101575, Article 101575. https://doi.org/10.1016/j.jengtecman.2020.101575

Kim, H. J., Kim, T. S., & Sohn, S. Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems*, *130*, 113229.

Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal*, *31*(6), 602-628. https://doi.org/https://doi.org/10.1002/smj.829

Mutlu, E. C., Oghaz, T., Rajabi, A., & Garibay, I. (2020). Review on Learning and Extracting Graph Features for Link Prediction [Review]. *Machine Learning and Knowledge Extraction*, *2*(4), 672-704. https://doi.org/10.3390/make2040036

Park, I., Jeong, Y., Yoon, B., & Mortara, L. (2015). Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. *Technology Analysis & Strategic Management*, *27*(7), 759-781. https://doi.org/10.1080/09537325.2014.971004

Park, M., & Geum, Y. (2022). Two-stage technology opportunity discovery for firm-level decision making: GCN-based link-prediction approach. *Technological Forecasting and Social Change*, *183*, 121934. https://doi.org/https://doi.org/10.1016/j.techfore.2022.121934

Qi, Y., Zhang, X., Hu, Z. Y., Xiang, B., Zhang, R., & Fang, S. (2022). Choosing the right collaboration partner for innovation: a framework based on topic analysis and link prediction. *Scientometrics*. https://doi.org/10.1007/s11192-022-04306-9

Ryan, M. (2020). The Future of Transportation: Ethical, Legal, Social and Economic Impacts of Self-driving Vehicles in the Year 2025. *Science and Engineering Ethics*, *26*(3), 1185-1208. https://doi.org/10.1007/s11948-019-00130-2

Schilling, M. A. (2012). *Strategic Management of Technological Innovation* (4th edition ed.). McGraw-Hill.

Shah, R. H., & Swaminathan, V. (2008). Factors influencing partner selection in strategic alliances: The moderating role of alliance context. *Strategic Management Journal*, *29*(5), 471-494. https://doi.org/10.1002/smj.656

Song, B., Seol, H., & Park, Y. (2016). A patent portfolio-based approach for assessing potential R&D partners: An application of the Shapley value. *Technological Forecasting and Social Change*, *103*, 156-165. https://doi.org/10.1016/j.techfore.2015.10.010

Vosooghi, R., Puchinger, J., Jankovic, M., & Vouillon, A. (2019). Shared autonomous vehicle simulation and service design. *Transportation Research Part C-Emerging Technologies*, *107*, 15-33. https://doi.org/10.1016/j.trc.2019.08.006

Wang, J. F., Wang, N. T., Zhao, W. Y., & Feng, L. J. (2025). Identifying and evaluating R&D partners via patent-based multilayer networks from the perspective of knowledge complementarity: A case study of unmanned ship technology. *Computers & Industrial Engineering*, *204*, Article 111085. https://doi.org/10.1016/j.cie.2025.111085

Wang, M.-Y. (2012). Exploring potential R&D collaborators with complementary technologies: the case of biosensors. *Technological Forecasting and Social Change*, *79*(5), 862-874.

Wang, X., Wang, Z., Huang, Y., Liu, Y., Zhang, J., Heng, X., & Zhu, D. (2017). Identifying R&D partners through Subject-Action-Object semantic analysis in a problem & solution pattern. *Technology Analysis & Strategic Management*, *29*(10), 1167-1180.

Yan, E., & Guns, R. (2014). Predicting and recommending collaborations: An author-, institution-, and country-level analysis. *Journal of Informetrics*, *8*(2), 295-309.

Yoon, B., & Song, B. (2014). A systematic approach of partner selection for open innovation. *Industrial Management & Data Systems*, *114*(7), 1068-1093.