

ERIA WORKING PAPER NO. 2

The Architecture of ASEAN's External Free Trade Agreements: A Textual Similarity Analysis

Yoshitaka Watanabe

The Architecture of ASEAN's External Free Trade Agreements: A Textual Similarity Analysis

Yoshitaka Watanabe

September 2026

As the Association of Southeast Asian Nations (ASEAN) and its member states have expanded their regional, plurilateral, and bilateral trade agreements, they have developed one of the world's most densely interconnected FTA networks. This overlapping architecture raises questions about where ASEAN's external FTAs converge around recurring textual models and where they remain fragmented. This paper applies a quantitative text-as-data approach to 65 FTAs across 12 substantive chapters. Using character 5-gram Jaccard similarity, hierarchical clustering, largest-cluster-share analysis, and party-level measures of internal and external similarity, it maps patterns of textual proximity and clustering.

The findings show that convergence is highly uneven across policy areas. Trade in services, investment, rules of origin, and trade remedies display relatively high similarity and form broad clusters at comparatively strict distance thresholds, a pattern consistent with established drafting structures and earlier treaty models. Trade in goods, customs procedures, sanitary and phytosanitary measures, and technical barriers to trade show intermediate convergence: detailed wording varies, but clusters expand rapidly as the threshold is relaxed, suggesting shared institutional frameworks and basic chapter structures. Competition, electronic commerce, intellectual property, and government procurement exhibit weaker or more uneven convergence, with field-specific configurations rather than a single common pattern.

Cluster analysis also identifies recurring partner-centred groupings, most pronounced for Japan and the United Arab Emirates, alongside more limited ASEAN-centred and mega-regional clustering. Party-level analysis further shows that Japan and the European Union combine relatively high internal similarity with relatively low external similarity, indicating coherent but comparatively distinctive treaty language. ASEAN, treated as a collective contracting party, lies near the sample means, consistent with the common structure of ASEAN+1 FTAs and their proximity to broader regional treaty practice.

Overall, ASEAN's external FTA architecture is layered rather than homogeneous. Broad convergence coexists with localised and party-specific patterns of clustering and fragmentation, while textual similarity alone does not establish substantive legal equivalence or the direction of treaty influence.

Keywords: ASEAN; Free Trade Agreements; Textual Similarity; Text-as-Data; Trade Policy; Regional Integration

1. Introduction

Amid the prolonged stagnation of multilateral trade negotiations under the World Trade Organization (WTO), countries have increasingly turned to regional, plurilateral, and bilateral trade agreements as alternative venues for advancing trade liberalisation and developing new rules (Allee et al., 2017b). While early free trade agreements (FTAs)¹ were primarily focused on tariff reductions and market access for goods, contemporary FTAs cover a much broader range of policy areas. These include trade in services, investment, intellectual property, competition, government procurement, electronic commerce, and, in some cases, environmental and labour standards. As a result, trade agreements have become not only instruments of market liberalisation but also vehicles for regulatory co-operation, institutional design, and the diffusion of legal disciplines across jurisdictions.

This transformation is particularly evident in Southeast Asia. Since the early 2000s, the member states of the Association of Southeast Asian Nations (ASEAN) have concluded numerous FTAs with partners across East Asia, Oceania, the Americas, Europe, and the Middle East (Kawai and Wignaraja, 2011). Beyond these individual efforts, ASEAN as a regional grouping has also developed a dense network of ASEAN+1 FTAs with major external partners. More recently, mega-regional FTAs such as the Regional Comprehensive Economic Partnership (RCEP) and the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) have further reshaped the region's trade governance architecture. Taken together, ASEAN and its member states now stand at the center of one of the world's most densely interconnected networks of FTAs.

At the same time, this dense treaty network has generated considerable complexity. Although RCEP has consolidated parts of the existing ASEAN+1 framework and created a common platform for regional trade rules, a 'noodle bowl' of overlapping bilateral, plurilateral, and regional FTAs remains (Asian Development Bank, 2022; Kawai and Wignaraja, 2011). These FTAs differ not only in their membership and economic coverage but also in the structure, language, and legal depth of their provisions. Consequently, it remains difficult to obtain a comprehensive picture of how disciplines in trade agreements have evolved across different policy areas, and to what extent they have converged around common templates or diverged according to the preferences of individual ASEAN Member States and negotiating partners.

Understanding this complexity is not merely an analytical exercise. Recent policy discussions have also highlighted the role of ASEAN's trade agreements within a broader strategy of regional resilience and agency in an increasingly fragmented geoeconomic environment. The ASEAN Geoeconomics Report 2025 argues that ASEAN

¹ The WTO generally uses the term 'regional trade agreements' (RTAs), whereas the broader term 'preferential trade agreements' (PTAs) is often used in the academic literature. Since this paper examines trade agreements involving ASEAN or its member states, it does not distinguish amongst RTAs, PTAs, and FTAs, and refers to them collectively as 'FTAs.'

should respond to external shocks through co-ordinated action, including the full and effective implementation of RCEP and the deepening of external partnerships through ASEAN Plus 1 FTAs and beyond (ASEAN Geoeconomics Task Force, 2025). It further frames such efforts as part of ASEAN's broader attempt to strengthen economic resilience and agency, reinforce centrality, and support open and inclusive rules-based trade. This policy context reinforces the importance of empirically mapping the structure of ASEAN's external FTAs. Before any assessment can be made of how ASEAN might use its treaty network strategically, it is necessary to understand where the underlying legal texts converge, where they remain fragmented, and which policy areas or treaty parties appear to anchor recurring drafting patterns.

Existing literature has shown that the design of trade agreements is not created in isolation. Major actors in global trade governance, such as the United States and the European Union, have developed distinctive approaches to treaty design and have played important roles in shaping the institutional design and substantive scope of trade agreements (Baccini et al., 2015). In addition, recent text-as-data research has demonstrated that many preferential trade agreements reproduce substantial portions of earlier treaty language, often through the use of models, templates, or prior agreements (Allee and Elsig, 2019). Taken together, these studies suggest that FTA texts and structures often reflect not only functional needs, but also the legal preferences, negotiating strategies, institutional traditions, and drafting practices of their key proponents. However, a central question remains insufficiently explored: whether similar patterns of model-based diffusion and textual convergence can be observed within ASEAN's external FTA network, or whether this network displays characteristics that differ from those of major global rule-making powers. More specifically, to what extent do ASEAN and its member states rely on common textual templates, and in which policy areas do ASEAN's external FTAs exhibit greater convergence or fragmentation?

Addressing this question is critical because ASEAN occupies a distinctive position in contemporary trade governance. ASEAN is neither a single negotiating actor comparable to the EU nor merely a collection of unrelated bilateral treaty-makers. Rather, its external FTA network combines agreements concluded collectively by ASEAN with individual FTAs concluded by its member states (Kleimann, 2013). This dual structure makes ASEAN an especially interesting case for examining how trade rules diffuse, cluster, and evolve across overlapping legal instruments. If ASEAN's external FTAs display strong textual convergence in certain fields, this may indicate the emergence of shared regional templates or the influence of dominant external partners. If, by contrast, they remain highly fragmented, this may suggest persistent diversity in negotiating strategies, regulatory preferences, or area-specific policy priorities amongst ASEAN Member States.

Against this backdrop, this paper seeks to map the structure of ASEAN's external FTAs through a quantitative analysis of treaty texts. Specifically, it adopts a text-as-data approach to measure textual similarity across FTAs and across major substantive areas.

By treating treaty provisions as data, this study aims to move beyond case-by-case comparison and provide a broader empirical picture of convergence and divergence within ASEAN's external FTA network.

The study offers three primary contributions. First, it provides a comprehensive quantitative overview of ASEAN's external FTAs across a wide range of substantive chapters. Second, it identifies policy areas in which treaty language appears to converge around common templates, as well as areas where provisions remain more diverse and fragmented. Third, it sheds light on the treaty practice of individual ASEAN Member States and their external partners by examining the degree to which their FTAs resemble one another. Through this analysis, the paper aims to clarify the structure of ASEAN's external FTA network and to offer a foundation for further research on the evolution of regional trade governance in Asia.

2. Overview of ASEAN's Trade Agreement Network

2.1 Scope of the Analysis

ASEAN's trade agreement network has developed in the broader context of its gradual approach to regional economic integration since the 1970s. Following the establishment of the ASEAN Free Trade Area (AFTA) in the 1990s, this process culminated in the launch of the ASEAN Economic Community (AEC) in 2015, which sought to advance a deeper and more comprehensive form of regional integration. Given the diversity of economic development levels amongst its member states, ASEAN has pursued liberalisation while maintaining a distinctive degree of flexibility, emphasising gradualism and consensus. This approach has also shaped ASEAN's external trade strategy.

In parallel with the deepening of intra-ASEAN integration, ASEAN and its member states have actively developed FTAs with external partners since the early 2000s. At the regional level, ASEAN has concluded six ASEAN+1 agreements with China, the Republic of Korea, Japan, Australia and New Zealand, India, and Hong Kong. These agreements have provided ASEAN with a platform for engaging major external partners collectively and have contributed to the formation of trade disciplines beyond ASEAN itself.

At the same time, ASEAN Member States have concluded a large number of bilateral and plurilateral FTAs with partners across East Asia, Oceania, the Americas, Europe, and the Middle East. Singapore, in particular, has developed an extensive network of FTAs. Other ASEAN Member States, including Malaysia, Viet Nam, Indonesia, Thailand, the Philippines, Cambodia, and Brunei Darussalam, have also concluded FTAs with external partners. As a result, ASEAN's external FTA network consists not only of agreements concluded by ASEAN as a whole, but also of a dense layer of agreements negotiated by individual member states.

This network has also been influenced by two landmark mega-regional agreements: the Regional Comprehensive Economic Partnership (RCEP) and the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP). RCEP includes ten ASEAN Member States and several of ASEAN's major external partners, while the CPTPP includes four ASEAN Member States. Stemming from the TPP initially spearheaded by the United States before its withdrawal, the CPTPP was designed to establish high-standard rules covering a wide range of advanced regulatory areas, including state-owned enterprises, labour and environment, and regulatory coherence. RCEP, by contrast, emerged from the existing ASEAN+1 FTA architecture and was negotiated with ASEAN playing a central role (Ravenhill, 2016; Rillo et al., 2022). Its final design reflects a more gradual and consensus-oriented approach, including flexibility for economies at different levels of development and the use of built-in agendas to promote further liberalisation over time (Asian Development Bank, 2022; Malingrey and Duval, 2022).

Against this backdrop, this paper focuses on ASEAN's external FTA network. Specifically, it examines FTAs concluded by ASEAN collectively, as well as those concluded individually by ASEAN Member States, with external partners. The dataset covers agreements signed by the end of 2025 for which official English texts were available for analysis. To maintain the focus on ASEAN's external trade relations, intra-ASEAN agreements such as ATIGA, ATISA, and ACIA are excluded.² To ensure analytical comparability, this study also excludes preferential trade arrangements and framework agreements with limited substantive coverage³, as well as stand-alone sector-specific agreements, such as investment agreements and digital partnership agreements.⁴ The resulting core dataset consists of 65 FTAs in total: six ASEAN+1 FTAs, two mega-regional FTAs (RCEP and CPTPP), and 57 FTAs concluded by individual ASEAN Member States (see Appendix Table A). The analysis reflects post-signature upgrades and amendments where the relevant official English texts are available and where they concern the chapters analysed in this study.⁵

² These refer to the ASEAN Trade in Goods Agreement (ATIGA), the ASEAN Trade in Services Agreement (ATISA), and the ASEAN Comprehensive Investment Agreement (ACIA), respectively.

³ Examples include the Lao PDR–Thailand Preferential Trading Agreement and the Framework Agreement on Comprehensive Economic Cooperation between India and Thailand. In addition, agreements concluded with the United Kingdom following its withdrawal from the European Union (Brexit) are excluded, as they primarily replicate and incorporate existing EU treaty provisions and do not possess independent structural frameworks.

⁴ In contrast, the Digital Economy Agreement between Australia and Singapore (DEA) and similar instruments with EFTA are included in the analysis as updates, given that they legally constitute subsequent amendments to pre-existing FTAs rather than stand-alone agreements.

⁵ For agreements concluded individually by ASEAN Member States, post-signature upgrades are concentrated mainly in Singapore's agreements. These include one-time or successive upgrades of FTAs with Japan, New Zealand, Australia, EFTA, India, and China, while the Indonesia–Japan EPA has also been upgraded. Amongst ASEAN+1 agreements, the agreements with Japan, Australia and New Zealand, and China have also been upgraded.

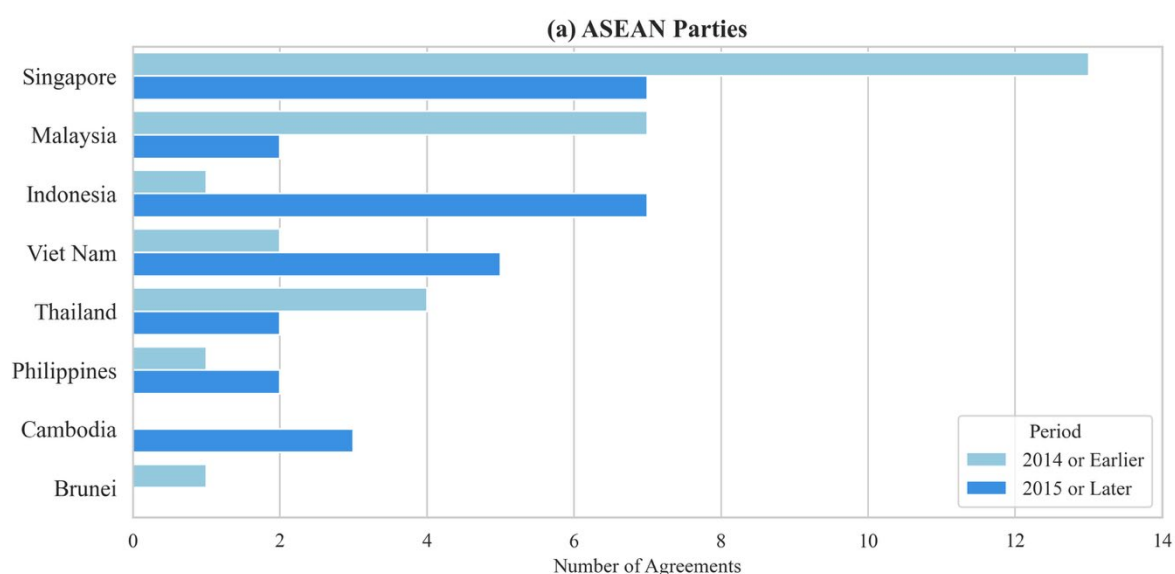
2.2 Trends in ASEAN Parties and External Partners in Individual FTAs

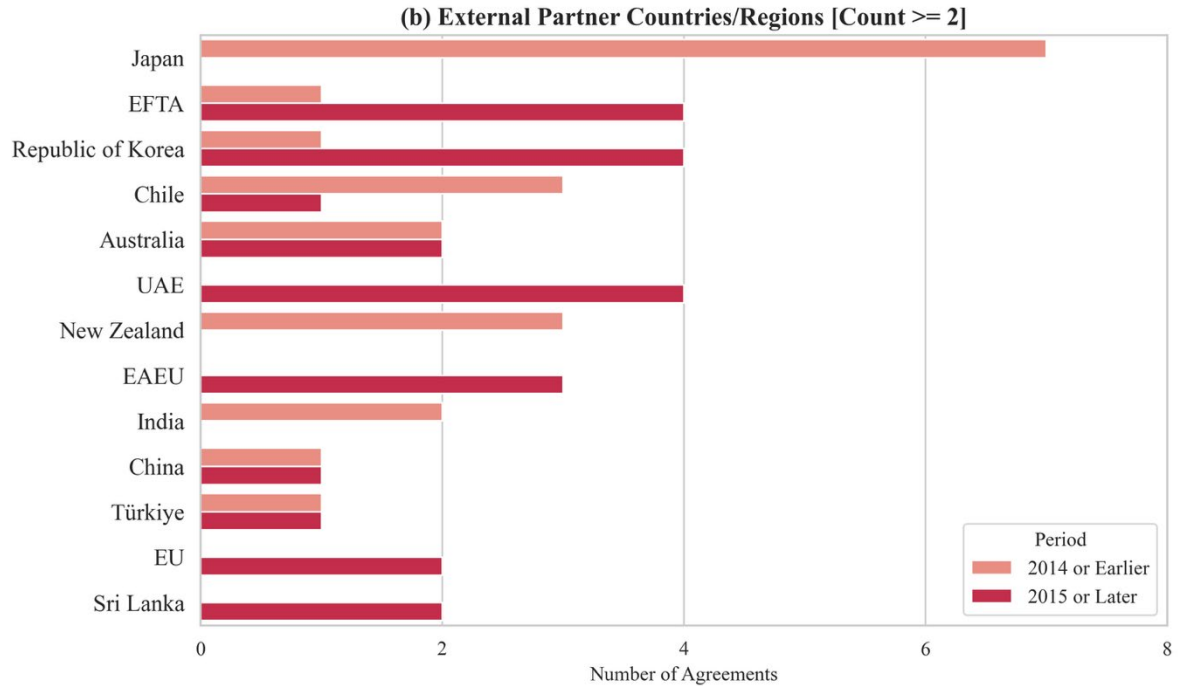
Figure 1 shows the distribution of ASEAN parties and major external partner countries and regions in FTAs covered by this study, excluding ASEAN+1 FTAs and mega-regional FTAs. The agreements are divided into two periods: those signed up to 2014 and those signed in or after 2015.

With respect to ASEAN parties, Singapore accounts for more than one-third of the FTAs in the full sample. This reflects Singapore's long-standing role as the most active ASEAN member state in concluding bilateral FTAs with external partners. However, the post-2015 period shows a gradual diversification amongst ASEAN parties. Indonesia and Viet Nam, in particular, increased the number of their external FTAs during this period, narrowing the gap with Singapore. This suggests that the ASEAN external FTA network has become more broadly distributed across ASEAN Member States over time.

The geographical distribution of external partners also underwent a significant shift between the two periods. Before 2015, individual FTAs by ASEAN Member States were mainly concluded with traditional economic partners such as Japan, Australia, Chile, and New Zealand. After 2015, however, the network expanded to include a wider range of partners and regions. The conclusion of multiple FTAs with the European Free Trade Association (EFTA), the European Union (EU), the Eurasian Economic Union (EAEU), the United Arab Emirates (UAE), the Republic of Korea, and Sri Lanka indicates that ASEAN Member States have increasingly diversified their external trade relationships beyond traditional Asia-Pacific frameworks.

Figure 1: ASEAN Parties and External Partners in Individual FTAs





Note: The figures cover only the individual FTAs included in the study dataset, namely agreements signed by the end of 2025 for which official English texts were available for analysis. In panel (b), only partner countries or regions involved in at least two FTAs in either period or in the full sample are shown. ASEAN+1 agreements, RCEP, and CPTPP are excluded from the counts.

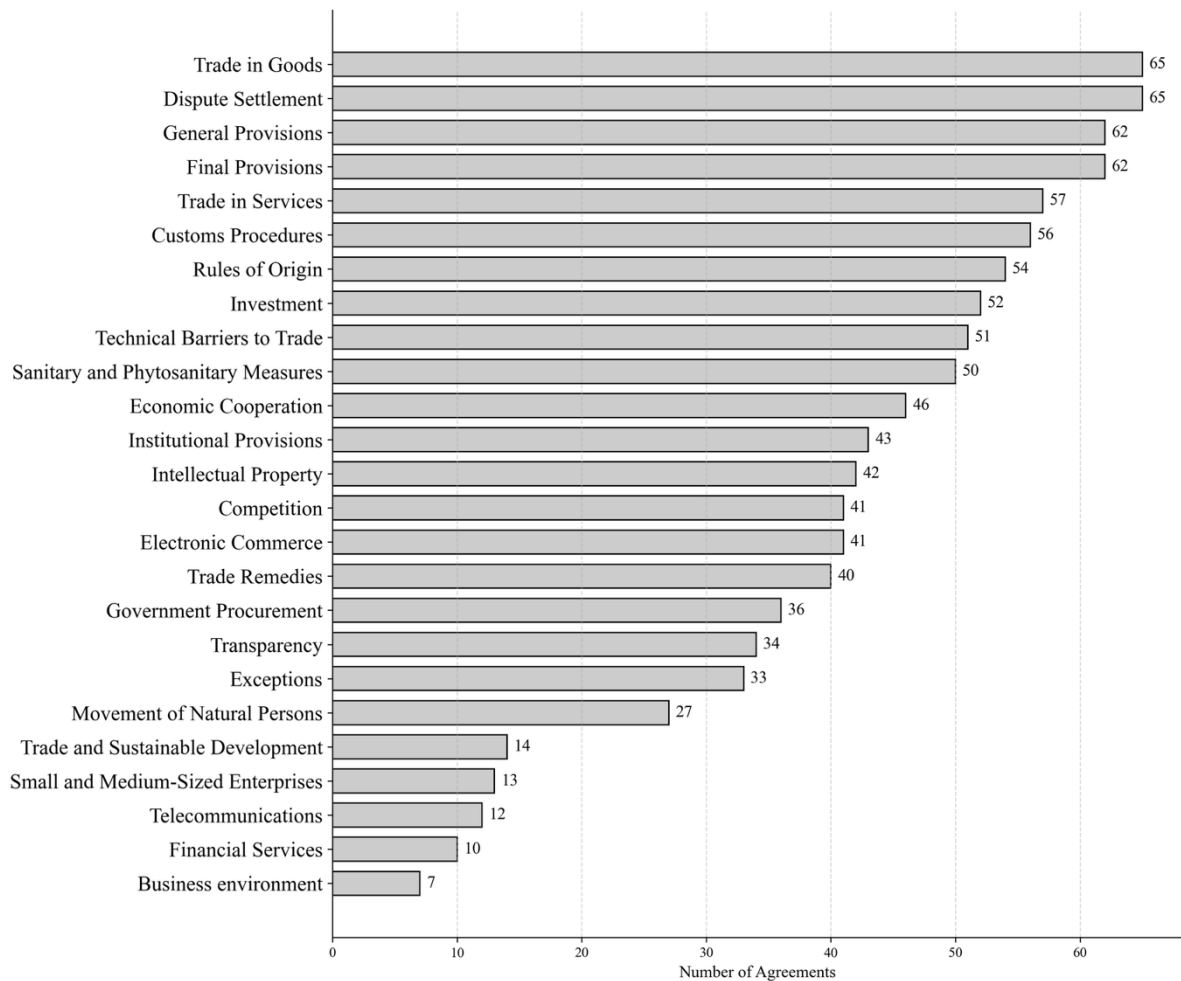
Source: Compiled by the author based on the Japan External Trade Organization (JETRO), World FTA Database (in Japanese) (accessed 1 May 2026).

2.3 Frequency of Major Chapters

Figure 2 shows the frequency of major chapters appearing in the 65 ASEAN external FTAs covered by this study. Because specific chapter titles vary across FTAs, they were standardised for counting purposes where they refer to closely related substantive fields. The labels in the figure use the most frequently appearing chapter titles.

The figure shows that more than half of the ASEAN external FTAs covered in this study contain not only chapters relating to trade in goods – such as rules of origin, customs procedures, sanitary and phytosanitary (SPS) measures, and technical barriers to trade (TBT) – but also chapters on trade in services, investment, intellectual property, electronic commerce, and government procurement. Although less frequent, more than ten FTAs also include chapters on small and medium-sized enterprises, as well as chapters addressing sustainability-related issues, including environment and labour, or environment alone.

Figure 2: Major Chapters Appearing in ASEAN External FTAs Covered in This Study



Note: Chapters appearing in fewer than five FTAs are excluded. For counting purposes, closely related chapter titles are standardised under common labels where appropriate. Where a single chapter covers multiple policy areas, such as 'Technical Regulations and Sanitary and Phytosanitary Measures' or 'Liberalisation of Investment, Trade in Services and Electronic Commerce,' it is counted under each relevant heading, provided that the chapter title explicitly refers to that policy area. For ASEAN–India and ASEAN–Korea, separate legal instruments on trade in goods, trade in services, investment, and dispute settlement mechanisms are treated as chapter-equivalent units.

Source: Compiled and calculated by the author.

Figure 2 should be understood as a formal count based primarily on independent chapter titles, and does not necessarily indicate the presence or absence of substantive disciplines in each policy area. Even where an agreement does not contain an independent chapter for a given field, relevant provisions may still be set out in other chapters or annexes. For example, provisions on rules of origin, SPS, TBT, trade remedies, general and security exceptions, and transparency are often placed in the trade in goods chapter where an agreement lacks independent chapters for these areas. Such structural differences reflect not only changes in drafting trends over time, but also the influence of particular FTA templates and drafting practices.

Several examples from the dataset illustrate this point. Amongst the FTAs covered in this study, those concluded by Japan do not include an independent chapter on trade remedies. Instead, provisions regarding anti-dumping measures, countervailing duties, and safeguards are included in the chapter on trade in goods. EFTA-related FTAs provide another example: rules on rules of origin and customs procedures are often placed in annexes rather than in independent chapters, while some chapters, such as those on intellectual property, may be relatively brief and supplemented by annexes.

These structural differences raise an important methodological question for the present analysis: how should the object of comparison be defined when measuring the similarity of FTA texts? If the purpose were to measure the overall similarity of FTA templates, it might be possible to compare each FTA as a single integrated text. However, to analyse similarities and differences in substantive disciplines by policy area, a more granular approach is required.

Previous studies analysing the similarity of trade agreements using text-as-data approaches have generally compared agreements as whole texts, while also conducting more disaggregated comparisons at the level of chapters or specific policy fields (Allee and Elsig, 2019; Allee and Lugg, 2016; Allee et al., 2017a; Allee et al., 2017b; Alschner et al., 2017; Frank, 2023). Lee and Tan (2025) also compare textual similarity at the level of corresponding articles. These studies indicate that the choice of unit of comparison is itself a methodological issue, particularly when the objective is to assess convergence or divergence in specific legal disciplines.⁶

This paper does not conduct manual legal coding or detailed semantic analysis. Instead, it uses a text-as-data approach to map patterns of textual similarity across ASEAN external FTAs, while seeking to compare substantive disciplines rather than merely formal treaty structures. For this reason, the analysis is conducted primarily at the chapter level, but with certain adjustments. Specifically, annexes or clearly identifiable sections are included in the comparison as chapter-equivalent units where they correspond to policy areas treated as independent chapters in other FTAs. Where no clearly identifiable textual unit exists, scattered provisions are not extracted and reconstructed into an independent chapter for the purpose of analysis. Conversely, where an FTA contains an independent chapter, it is not excluded from the analysis merely because that chapter consists only of very short provisions. In the following analysis, the term ‘chapter’ is therefore used as shorthand to refer to independent chapters and, where appropriate, chapter-equivalent textual units such as annexes or clearly identifiable sections.

⁶ Although not based on textual similarity, related projects have used expert assessment to transform treaty provisions into structured data on legal disciplines and institutional features. Representative examples include the World Bank’s Deep Trade Agreements (DTA) database and the Design of Trade Agreements (DESTA) project. See Mattoo et al. (2020); Dür et al. (2014).

Based on this approach, the following textual analysis focuses on twelve substantive chapters that appear in more than half of the FTAs in the dataset, excluding general, institutional, and procedural chapters such as General Provisions, Institutional Provisions and Dispute Settlement. These twelve chapters are: trade in goods, rules of origin, customs procedures, sanitary and phytosanitary measures (SPS), technical barriers to trade (TBT), trade remedies, trade in services, investment, intellectual property, competition, electronic commerce, and government procurement. In addition, each of these chapters appears in FTAs concluded by at least six individual ASEAN Member States, in at least one ASEAN+1 FTA, and in both CPTPP and RCEP.

3. Analytical Methodology

As explained in the previous section, the analysis that follows focuses on twelve substantive chapters that appear in more than half of the agreements in the dataset. The purpose is to compare the textual similarity of agreement provisions across policy areas and to examine whether ASEAN's external FTAs exhibit patterns of convergence, divergence, or clustering. Before setting out the specific method used to measure textual similarity, this section first clarifies what high textual similarity may imply in the context of trade agreement analysis.

3.1 Implications of Textual Similarity

What does high textual similarity between international economic agreements actually imply? In academic writing or literary works, the close replication of existing text may raise concerns about plagiarism or copyright infringement. In legal drafting, however, textual re-use is not necessarily problematic. On the contrary, the use of established formulations, standard clauses, and so-called boilerplate language is a common and often desirable practice in contracts, legislation, and treaty drafting. Similar practices have also been observed in the drafting of international economic agreements, where treaty texts often incorporate language from earlier agreements or established templates (Allee and Elsig, 2019; Alschner et al., 2017).

This continuity offers significant legal and practical value, as legal texts are expected to produce stable and predictable effects. Drafters often prefer language whose meaning has already been tested in practice, interpreted by institutions, or accepted by negotiating parties. Introducing novel wording may inadvertently create uncertainty, even where no substantive change is intended. In the context of international trade, the reuse of existing treaty language, including language drawn from established WTO agreements, may help maintain consistency across agreements, reduce negotiation costs, and avoid unintended legal consequences (Allee et al., 2017b).

Previous text-as-data studies have shown that such textual continuity may also reveal broader institutional and political dynamics. Treaty texts do not circulate randomly; they

are shaped by existing templates, negotiating practices, and the relative bargaining power of the parties. Repeated reproduction of particular treaty formulations may indicate the diffusion of a treaty model or the influence of a powerful state projecting its preferred language into new institutional settings. For example, studies have shown that the text of the Trans-Pacific Partnership (TPP) was disproportionately similar to earlier United States trade agreements, and that its investment chapter was heavily influenced by prior United States treaty practice (Allee and Lugg, 2016; Alschner and Skougarevskiy, 2016a). Treaty language may also diffuse indirectly through existing regional or mega-regional reference templates, including cases where the influence of a non-party is transmitted through prior agreements, as Frank (2023) shows in relation to RCEP.

Textual analysis can also help identify areas of relative legal innovation and change in treaty design. By comparing agreements over time, previous studies have examined where new agreements depart from earlier templates, combine existing formulations in new ways, or introduce new textual models. In the investment treaty context, Alschner and Skougarevskiy (2016b) use textual similarity to trace changes in national BIT models over time and broader patterns of treaty design diffusion across countries. In the FTA context, Allee et al. (2017a) find that the EU–Canada Comprehensive Economic and Trade Agreement (CETA) was less heavily drawn from previous EU and Canadian agreements than might have been expected. At the provision level, similar dynamics have been observed in preferential trade agreements: some provisions are repeatedly reproduced across agreements, while others remain isolated or diffuse only weakly, suggesting that treaty-language diffusion may vary substantially across policy areas and legal issues (Morin et al., 2019).

Building on these insights, this paper applies text-as-data techniques to a more focused dataset: ASEAN's external FTA network. Because the dataset is limited to FTAs involving ASEAN Member States and their external partners, the analysis does not attempt to trace the global origins of every provision or determine which external actor has exercised the greatest influence on treaty drafting. Instead, it examines whether, and in which policy areas, ASEAN's external FTAs display common textual patterns. By comparing FTA texts at the chapter level, the paper seeks to identify areas where treaty language has converged around common models and areas where agreement texts remain more fragmented.

This approach also makes it possible to examine differences across ASEAN Member States and external partners. Specifically, the paper uses textual similarity to explore whether individual ASEAN Member States tend to reproduce similar language across their own FTAs, or whether their agreements are more strongly shaped by the templates of their negotiating partners. To this end, the analysis introduces two related measures: internal similarity and external similarity (see Section 5).

3.2 Measuring Textual Similarity

To compare the textual similarity of FTA texts quantitatively, this study follows the text-as-data approach developed by Alschner and Skougarevskiy (2016b) and applied to preferential trade agreements by Alschner et al. (2017). Specifically, it calculates Jaccard similarity based on character 5-grams. In this approach, each agreement text is divided into consecutive sequences of five characters, and the resulting set of character strings is compared across FTAs. For example, the phrase 'each party' is segmented into overlapping five-character units, including spaces, such as 'each_', 'ach_p,' and 'ch_pa.' The Jaccard similarity between two texts is then calculated as the proportion of unique 5-grams shared by the two texts relative to the total number of unique 5-grams appearing in either text. The value ranges from 0 to 1, where 0 indicates that the two texts share no character 5-grams and 1 indicates that they are identical.

Before calculating Jaccard similarity, the treaty texts were cleaned and preprocessed to reduce the influence of editorial and formatting differences across FTAs. This step is necessary because the purpose of the analysis is to compare recurring treaty language and drafting patterns rather than document-specific formatting styles. Elements such as page numbers, footnotes, chapter, section, and article numbers and labels, list markers, and cross-reference markers may vary according to source documents, drafting conventions, and editing practices, and may therefore introduce noise into the analysis. These elements were removed to the extent possible. Hyphenated line breaks were corrected, and hyphens were removed so that minor spelling and formatting differences would not unduly affect similarity scores. The texts were then converted to lowercase, numbers, punctuation, and symbols were removed, and extra spaces were normalised.

The use of character 5-gram-based Jaccard similarity differs from Bag-of-Words approaches that rely primarily on the frequency of individual words. Because 5-grams preserve short sequences of characters, this method retains, to some extent, information about word order and formulaic legal phrasing. This feature is particularly useful for analysing trade agreements, which often rely on standardised expressions, recurring clauses, and treaty-specific drafting formulas. By comparing continuous character patterns, the method can capture not only the presence of common keywords but also the structural proximity of legal language. Character 5-grams are also relatively robust to minor variations in spelling, singular and plural forms, and small differences in expression, making them useful for detecting provisions that are not formally identical but are textually close.

Compared with large language models or more advanced machine learning techniques, this method is relatively simple. However, its simplicity is also an advantage. It avoids the 'black-box' problem inherent in more complex computational tools and makes the analysis more transparent and replicable. Since the similarity score is based on clearly

defined textual units and a straightforward set comparison, the results can be more easily interpreted and reproduced than those obtained from more complex models.

At the same time, Jaccard similarity should be understood as a measure of formal textual similarity. It does not assess legal similarity in a substantive or doctrinal sense. A high similarity score indicates that two texts share many character-level patterns, but it does not necessarily mean that their legal effects are identical. Conversely, even where two chapter texts have a similarity score of 0.90, the remaining 0.10 may include wording that is decisive for the scope of obligations or the legal effects of the provisions. This limitation is especially important in legal analysis, where small textual differences can have substantial normative consequences.

Nevertheless, modern FTAs, including ASEAN's external FTAs examined in this paper, share a certain degree of standardised chapter structure and legal drafting. For this reason, character 5-gram-based Jaccard similarity provides a useful tool for quantitatively capturing textual proximity amongst the FTAs. It enables the identification of groups of FTAs that share similar treaty language, as well as patterns associated with particular countries, partners, time periods, or policy areas.

4. Chapter-Level Similarity Patterns

Following the methodology set out in Section 3, this section examines chapter-level patterns of textual similarity across the 65 FTAs covered in this study. It calculates pairwise Jaccard similarity scores based on character 5-grams for each of the 12 chapters and analyses the results in three steps. First, Section 4.1 examines the overall level and dispersion of textual similarity within and across chapters. Second, Section 4.2 applies hierarchical clustering to identify groups of FTAs that exhibit relatively high textual similarity. Third, Section 4.3 uses the largest cluster share to examine whether textual similarity is confined to a small number of FTA pairs or is structured into broader clusters of agreements.

4.1 Distribution of Textual Similarity by Chapter

Figure 3 shows the distribution of pairwise Jaccard similarity scores across the 12 chapters, arranged in descending order of median similarity. The boxes represent the interquartile range, that is, the middle 50 per cent of pairwise similarity scores for each chapter. Although the results are limited to ASEAN's external FTAs covered in this study, the figure indicates that levels of textual similarity differ substantially across policy areas.

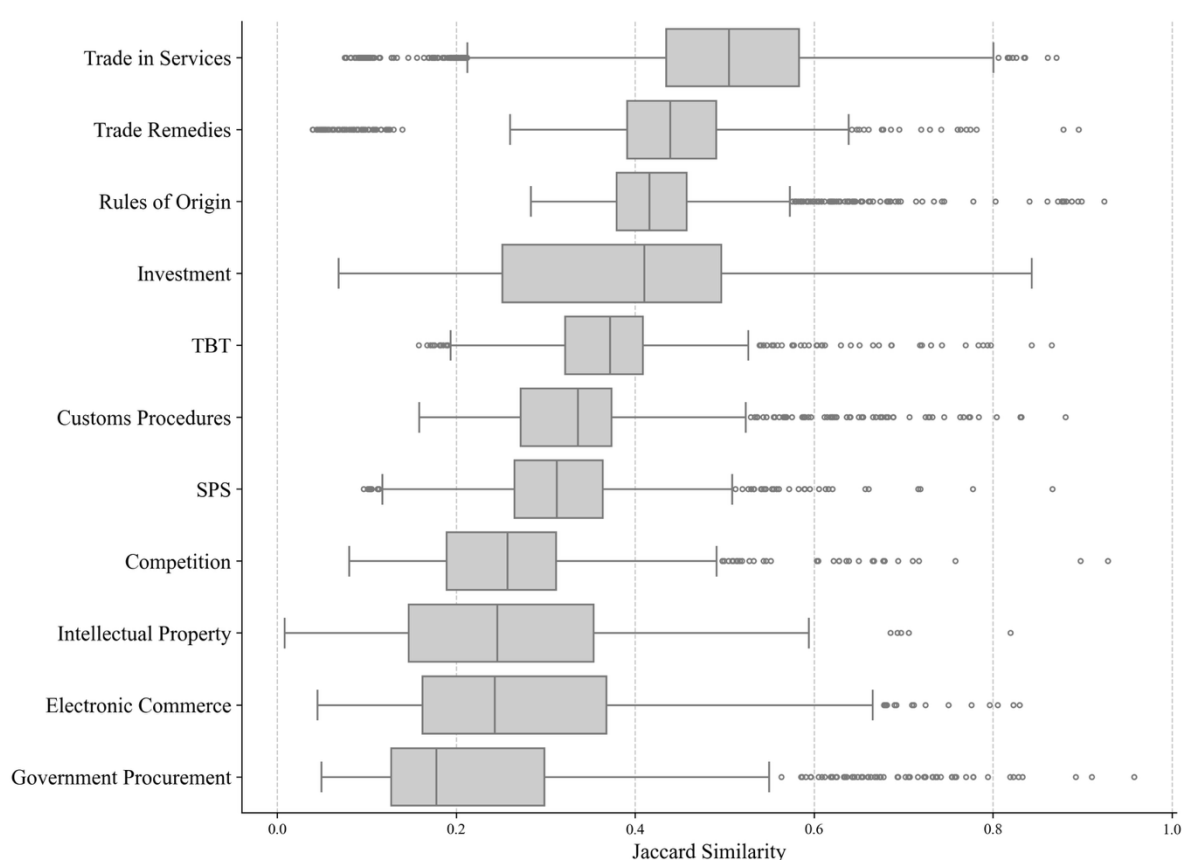
The chapters with relatively high median similarity are trade in services, trade remedies, investment, and rules of origin. In these areas, FTA texts appear to rely to a considerable extent on common article structures and recurring legal formulations. This suggests

that, within the dataset analysed in this paper, these chapters are more likely to be drafted on the basis of established templates or standard treaty language.

By contrast, competition, intellectual property, electronic commerce, and government procurement show lower median similarity and greater textual variation across FTAs. This may reflect differences in the depth of commitments, the degree of institutional maturity in each policy area, or the extent to which negotiating parties have developed and maintained their own preferred drafting models. Trade in goods, customs procedures, SPS, and TBT occupy an intermediate position, with median similarity scores falling between these two groups.

It is also important to note that even chapters with relatively low overall similarity often contain pairwise scores in the upper range of the distribution. This indicates that, although a chapter may be diverse as a whole, highly similar texts may still exist amongst particular FTA pairs or groups of agreements. In other words, a low median similarity does not necessarily imply the absence of templates or clusters. Rather, it may indicate that textual similarity is concentrated within specific subsets of FTAs. This point is examined in greater detail in the next section through hierarchical clustering.

Figure 3: Distribution of Pairwise Jaccard Similarity across the 12 Chapters



Source: Compiled and calculated by the author.

4.2 Cluster Formation by Chapter

This section applies hierarchical clustering to the pairwise Jaccard similarity data to identify groups of FTAs with relatively high textual similarity. Hierarchical clustering successively merges agreements that are close to one another until all agreements are ultimately integrated into a single cluster. In this analysis, Jaccard distance is defined as one minus the Jaccard similarity score, and the results are presented as dendrograms. To measure the distance between clusters, this paper uses the group-average method, under which the distance between two clusters is calculated as the average pairwise distance between the agreements contained in them.⁷

Figure 4 presents illustrative dendrograms for three chapters: trade in goods, trade in services, and electronic commerce. Each leaf represents an FTA and is labelled in the format 'ASEAN party–partner country or region.' The horizontal order of the leaves is generated automatically by the plotting algorithm and should not be interpreted as a substantive ranking. The height at which two FTAs or clusters merge represents their pairwise Jaccard distance, or the average pairwise distance between clusters under the group-average method. Accordingly, FTAs or clusters that merge at lower levels are more textually similar, whereas those that merge only at higher levels are more distant.

A cross-sectoral reading of the 12 dendrograms suggests three broad tendencies, although the frequency with which particular parties appear varies across chapters. These patterns are also broadly consistent with the cross-sectoral average similarity scores calculated for each FTA pair (see Appendix Table B for the pairs with the highest average similarities). First, a recurrent pattern is the formation of initial clusters around an external partner country or region. Amongst the major external partners that have concluded four or more bilateral FTAs with ASEAN Member States, such partner-centred initial clustering is visible across most or all comparable chapters for Japan, EFTA, Chile, and the UAE.⁸

At the same time, the strength and completeness of partner-centred clustering vary considerably across external partners. The tendency is most pronounced for Japan and the UAE: in more than half of the chapters analysed, all or nearly all of their relevant agreements form a single cohesive initial cluster. By contrast, initial clustering amongst agreements involving the Republic of Korea appears in fewer than half of the chapters. Australia shows partner-centred clustering only in exceptional cases. Instead, in more than half of the chapters, its agreements form localised clusters involving Australia and New Zealand and a common ASEAN party, such as Thailand–Australia with Thailand–

⁷ The group average method measures the distance between two clusters by taking the average distance across all pairs of agreements contained in the two clusters. It is useful for identifying groups of FTAs that are coherent as a whole. An alternative method is single linkage, which merges clusters on the basis of the closest pair of agreements between them. Single linkage tends to produce earlier merging and broader clusters because of the chaining effect. This paper therefore uses the group average method as the main specification and single linkage as a robustness check.

⁸ For this purpose, 'comparable chapters' refers to chapters available for comparison in at least two FTAs involving the relevant external partner.

New Zealand, and Malaysia–Australia with Malaysia–New Zealand.

Partner-centred clustering is not, however, uniform across all agreements involving the same external partner. In several chapters, multiple agreements form a relatively coherent initial cluster, while one or more agreements involving the same partner occupy separate positions in the dendrogram. In the trade in services dendrogram presented in Figure 4, for example, the Singapore–Japan FTA lies outside the principal cluster formed by Japan’s other bilateral FTAs, while the Philippines–EFTA FTA lies outside the grouping formed by several other EFTA agreements. Such differences may reflect the timing of the agreements, the ASEAN party involved, the institutional architecture adopted in the relevant chapter, subsequent upgrades, or reliance on common reference agreements. Both early agreements that lie outside clusters formed by later agreements and recently concluded or upgraded agreements that lie outside earlier groupings can be observed. The results therefore do not indicate a simple or unidirectional process of template evolution. Although they are consistent with the influence of drafting models associated with particular external partners, explaining individual clusters and outliers would require further chapter-level examination of the relevant treaty provisions and negotiating contexts.

Second, although less prevalent, several chapters contain initial clusters that are not organised primarily around a common external partner. The most recurrent pattern is clustering amongst agreements concluded by the same ASEAN party, particularly Singapore. Singapore-centred groupings appear in 11 of the 12 chapters analysed, although their composition varies across fields. Their frequent appearance should be interpreted in light of Singapore’s large share of the dataset: its 20 agreements account for nearly one-third of the 65 FTAs analysed. Recurring examples include groupings involving Singapore’s agreements with the EU, Türkiye, and Sri Lanka, as well as clusters linking its agreements with Peru, Costa Rica, Panama, and Chinese Taipei. Depending on the chapter, these groupings also include agreements with partners such as the United States, the Republic of Korea, the GCC, India, or MERCOSUR. Comparable ASEAN-party-centred clustering appears more selectively for Malaysia, with groupings visible in trade in goods and trade remedies. Such patterns are less frequent and less consistent across chapters than the Singapore-centred groupings.

The recurrence of Singapore-centred clusters may partly reflect Singapore-specific drafting practices. At the same time, the changing composition of these clusters suggests that some may also be associated with particular negotiating contexts, regional agreement networks, or shared reference texts. For example, Singapore’s agreements with Türkiye and Sri Lanka form close links in several chapters, and EU-related FTAs often appear in the same or adjacent subclusters. This is consistent with the possibility that the Singapore–EU FTA, or broader EU-related treaty language, was available as a reference point in those negotiations. Similarly, the proximity in several chapters amongst Singapore’s agreements with the United States, Peru, Costa Rica, Panama, and Chinese Taipei may reflect common trans-Pacific or US-related drafting influences.

However, because the dataset contains only one FTA involving the United States, it does not permit the identification of a broader US-centred pattern.

A separate and more limited pattern involves clustering around ASEAN collectively. In trade in goods, trade in services, and investment, several ASEAN+1 agreements form relatively coherent initial clusters. In all three chapters, however, the ASEAN–Australia–New Zealand FTA lies outside the principal ASEAN+1 grouping. These results suggest that ASEAN-centred drafting continuity may be stronger in particular fields, but does not extend uniformly across ASEAN's external agreements.

Third, several clusters show textual proximity to mega-regional FTAs. RCEP appears close in a number of chapters to several FTAs involving the Republic of Korea, as well as to the ASEAN–Australia–New Zealand FTA and the ASEAN–China FTA in the versions reflecting their respective upgrade protocols signed in 2023 and 2025. Given the position of these agreements within the broader East Asian FTA network, their proximity to RCEP may reflect the circulation of legal language and negotiating experience across related regional negotiations.

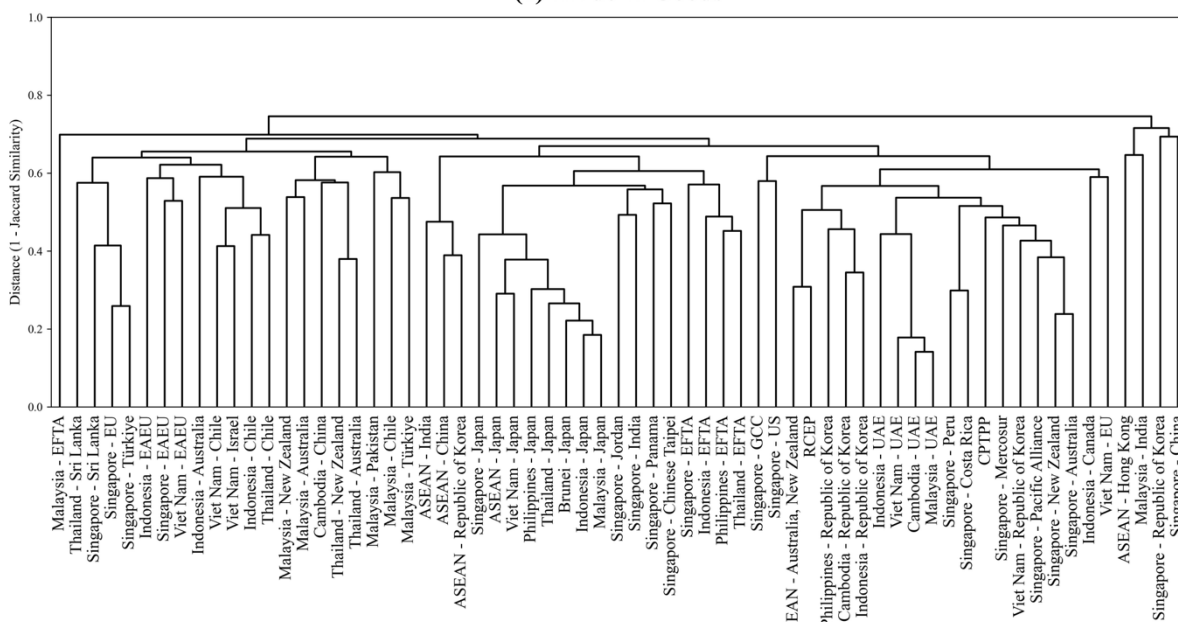
CPTPP displays a somewhat different pattern. It frequently appears close to the Singapore–Pacific Alliance FTA and, in some chapters, to FTAs involving Australia or New Zealand. The Indonesia–Canada FTA also appears close to CPTPP in several chapters, suggesting that CPTPP-related language may have served as a reference point in its negotiation. More broadly, these patterns suggest that drafting models associated with CPTPP may have circulated through agreement networks connected with Asia-Pacific and trans-Pacific rulemaking.

The dendrograms can also be interpreted in light of the substantive and institutional architecture of individual chapters. As noted above, Jaccard similarity measures formal overlap in textual expressions rather than substantive legal similarity as such, and its interpretation therefore requires caution. Nevertheless, legal content, drafting structure, and treaty language are closely related in FTAs, and clusters based on textual similarity often correspond to recognisable regulatory or drafting models. The trade in services dendrogram provides a clear example.

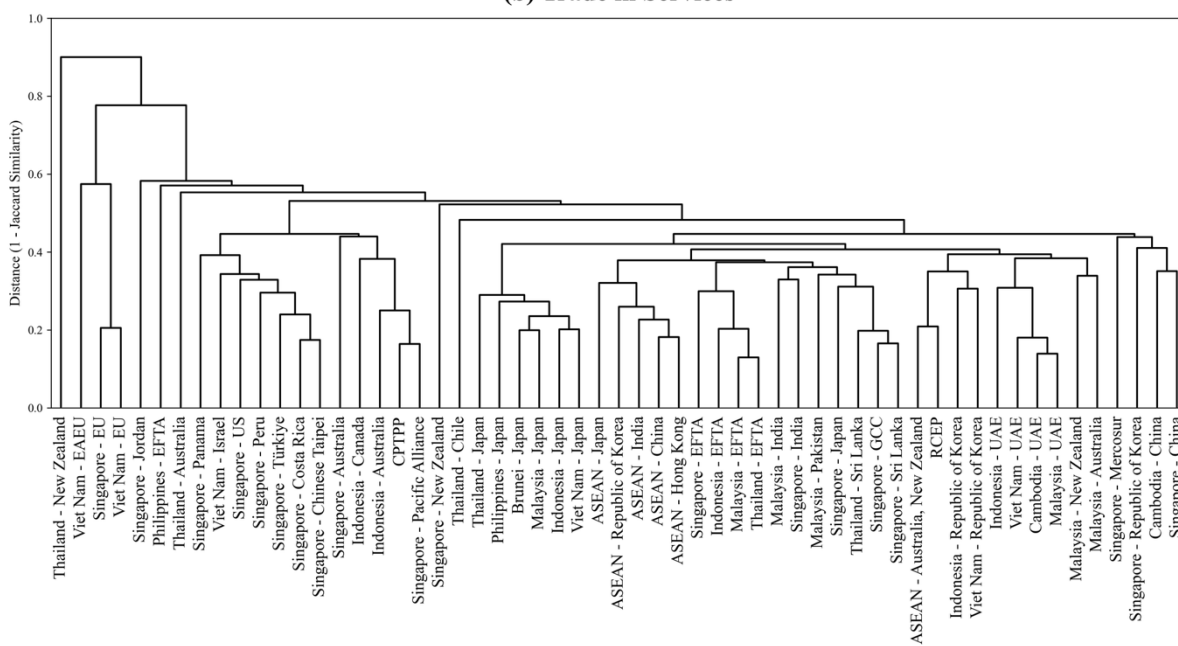
Services chapters differ not only in the substance of their commitments but also in the basic architecture through which those commitments are expressed. Positive-list FTAs generally follow a GATS-style scheduling model, under which obligations such as national treatment apply only to sectors inscribed in the relevant schedule and are subject to the terms, limitations, conditions, and qualifications set out therein (Serafica and Ramli, 2022). This produces relatively distinctive treaty language, including recurring references to scheduled sectors, modes of supply, and limitations on commitments. By contrast, negative-list FTAs generally formulate national treatment and most-favoured-nation treatment as broadly applicable obligations, while managing exceptions through lists of non-conforming measures.

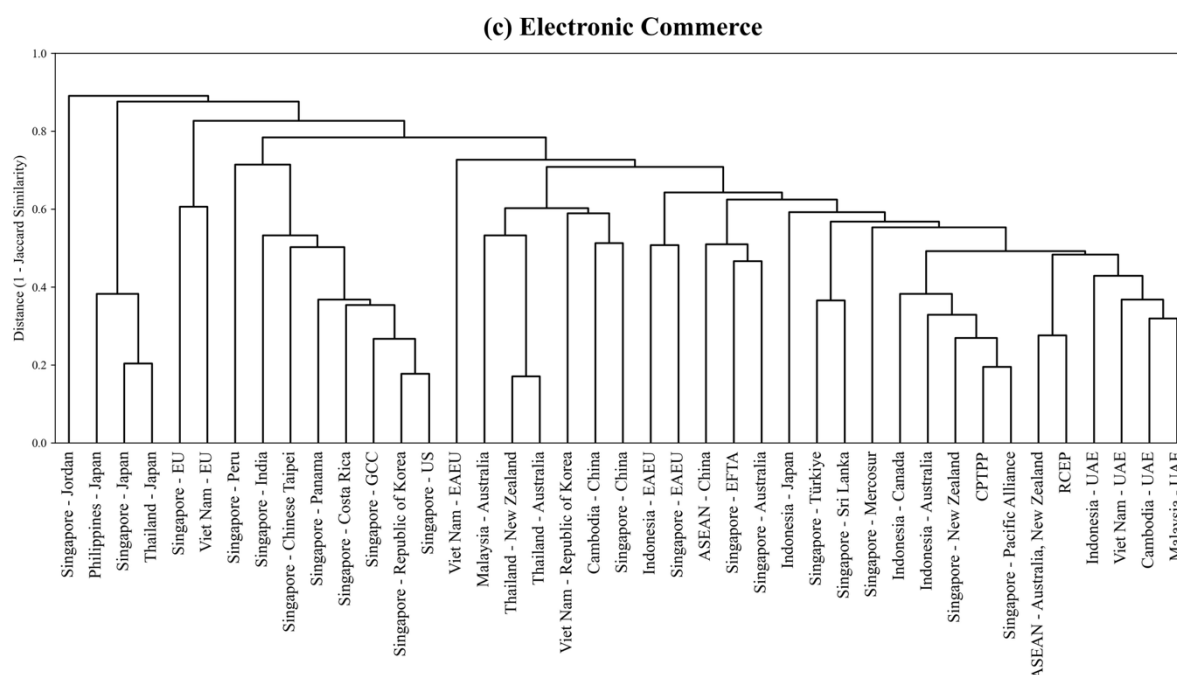
Figure 4: Examples of Dendrograms
(Trade in Goods, Trade in Services, and Electronic Commerce)

(a) Trade in Goods



(b) Trade in Services





Source: Compiled and calculated by the author.

Of the 57 FTAs in the dataset that contain a services chapter, 14 adopt a negative-list approach, excluding transitional or hybrid models such as RCEP. With limited exceptions, these FTAs use chapter titles such as ‘Cross-Border Trade in Services.’ They also tend to exclude Mode 3, or commercial presence, from the services chapter and regulate it primarily through the investment chapter, treating commercial presence as a form of establishment covering both services and non-services sectors. These architectural choices are reflected in the wording of the chapters. Negative-list FTAs often use shorter and more general formulations of national treatment and, where included, most-favoured-nation treatment, whereas positive-list FTAs more frequently retain GATS-style references to scheduled sectors and conditions. As a result, many negative-list agreements form a relatively coherent cluster in the services dendrogram, although the clustering is not complete because textual similarity also reflects party-specific drafting practices and other differences in chapter design.

Electronic commerce provides another useful example. In this field, the clustering pattern is broadly consistent with the chronological evolution of treaty practice. Earlier frameworks typically focus narrowly on paperless trading, addressing the electronic submission and exchange of trade-related administrative documents in order to reduce administrative costs. Subsequent FTAs expanded this approach into more comprehensive electronic commerce chapters, introducing standard rules such as the non-imposition of customs duties on electronic transmissions, online consumer protection, and personal data protection. More recent FTAs demonstrate a further shift towards broader and more advanced digital trade disciplines, with some agreements framing these rules under chapter titles such as ‘Digital Trade’ or the broader ‘Digital

Economy.’ These chapters may include more advanced provisions on cross-border data flows, data localisation requirements, source code, and co-operation on emerging technologies (Asian Development Bank, 2022).

These evolutionary stages are reflected in the clustering pattern. FTAs sharing similar conceptual scopes and terminology tend to form identifiable subclusters, corresponding to the gradual deepening of digital rulemaking in ASEAN’s external FTAs. Taken together with the results for trade in services, this suggests that the clustering patterns capture not only incidental similarities in wording but also broader choices in chapter architecture and regulatory design.

Overall, these patterns suggest that textual similarity in ASEAN’s external FTAs does not arise from a single source. It may reflect party-specific drafting models, changes over time and through subsequent upgrades, shared chapter architectures, context-specific reference points, regional and interregional agreement networks, and broader mega-regional rulemaking processes. The relative importance of these influences is likely to vary across chapters.

4.3 Changes in Cluster Structure across Distance Thresholds

The dendrograms show the distance at which individual FTAs or groups of FTAs merge with one another. They therefore provide useful information on how quickly small clusters, including those centred on particular countries, regions, or treaty models, become connected to larger groups. However, dendrograms are not always suitable for comparing all twelve chapters in a concise manner. To make the cross-sectoral comparison clearer, this section examines changes in the share of FTAs belonging to the largest cluster.

The largest cluster share is defined as the number of FTAs included in the largest cluster at a given distance threshold divided by the total number of FTAs in the relevant chapter. As the distance threshold becomes larger, FTAs with lower levels of similarity are allowed to merge, and the largest cluster generally expands. Thus, the speed at which the largest cluster share increases provides an indication of how quickly a field forms a broad textual cluster. A rapid increase at a low distance threshold suggests strong textual convergence, whereas a slow or limited increase suggests greater textual fragmentation.

Figure 5 shows the trajectory of the largest cluster share as the Jaccard distance threshold is gradually increased. Since the horizontal axis represents Jaccard distance, moving to the right means that the clustering criterion is relaxed. The figure groups the twelve chapters into three broad types according to the speed and extent of cluster expansion: an early convergence group, an intermediate convergence group, and a weak or uneven convergence group.

The first type consists of chapters in which the largest cluster share rises rapidly even under relatively strict or moderate distance thresholds. Trade in services, investment, rules of origin, and trade remedies fall into this group. In these areas, the largest cluster already accounts for a substantial share of FTAs at relatively low or moderate distance thresholds, and it approaches or exceeds a very high share as the threshold is relaxed. This indicates that textual similarity in these fields is not limited to isolated pairs of FTAs, but appears as a relatively large and coherent group of FTAs. These fields may therefore be characterised as areas of early convergence, where standardised treaty structures or widely shared legal formulations exert a relatively strong influence.

This interpretation is consistent with the boxplot analysis in Figure 3. These four areas have relatively high median similarity scores, and Figure 5 further shows that such similarity is not merely dispersed across individual FTA pairs. Rather, it is reflected in the formation of large clusters. In other words, Trade in services, investment, rules of origin, and trade remedies are not only similar on average; their similarities are reflected in broad clusters of FTAs sharing similar treaty language.

The second type consists of chapters in which the largest cluster share remains relatively low under strict thresholds but rises sharply when the distance threshold is increased. Trade in goods, customs procedures, SPS, and TBT fall into this group. In these areas, textual convergence is less apparent at the level of highly specific wording. However, once a broader standard of similarity is applied, a large proportion of FTAs become part of the same cluster.

This pattern suggests that these fields retain textual differences at the detailed drafting level, while sharing broader institutional structures or regulatory frameworks. For example, trade in goods, customs procedures, SPS, and TBT chapters are often shaped by multilateral trade law concepts and WTO-related disciplines, but individual FTAs may vary in the way they incorporate, elaborate, or supplement those disciplines. These areas may therefore be described as fields of intermediate convergence: they do not display early and strong convergence in precise wording, but they do show broad convergence once the comparison allows for greater textual distance.

The third type consists of chapters in which the formation of large clusters is slower, more limited, or more uneven. Competition, electronic commerce, intellectual property, and government procurement belong to this group. These areas are located toward the lower end of the similarity distribution in Figure 3, and Figure 5 confirms that they do not form large clusters as quickly or as extensively as the other fields. This indicates that textual convergence remains weaker or more uneven in these areas.

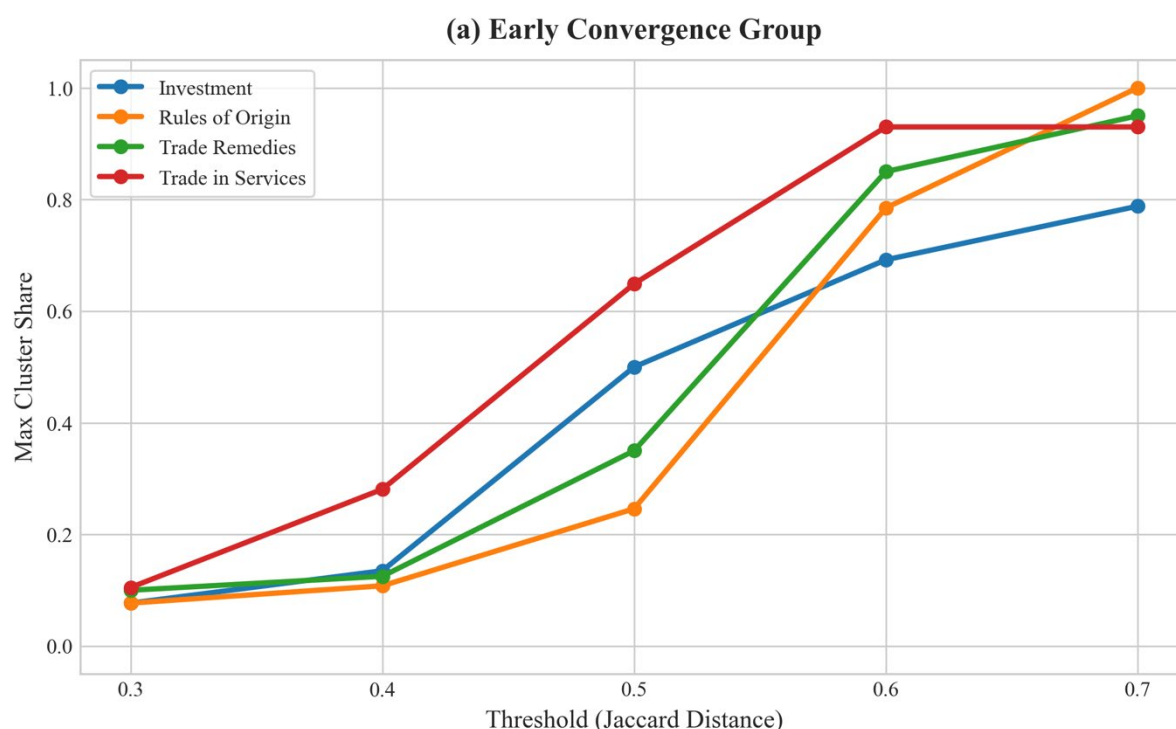
Amongst these four fields, however, the patterns are not identical. Competition shows a limited increase in the largest cluster share, suggesting that many FTAs remain textually separate even as the distance threshold is relaxed. Government procurement shows a somewhat larger cluster at lower thresholds than competition, but the expansion of its largest cluster is limited thereafter. This suggests that government procurement

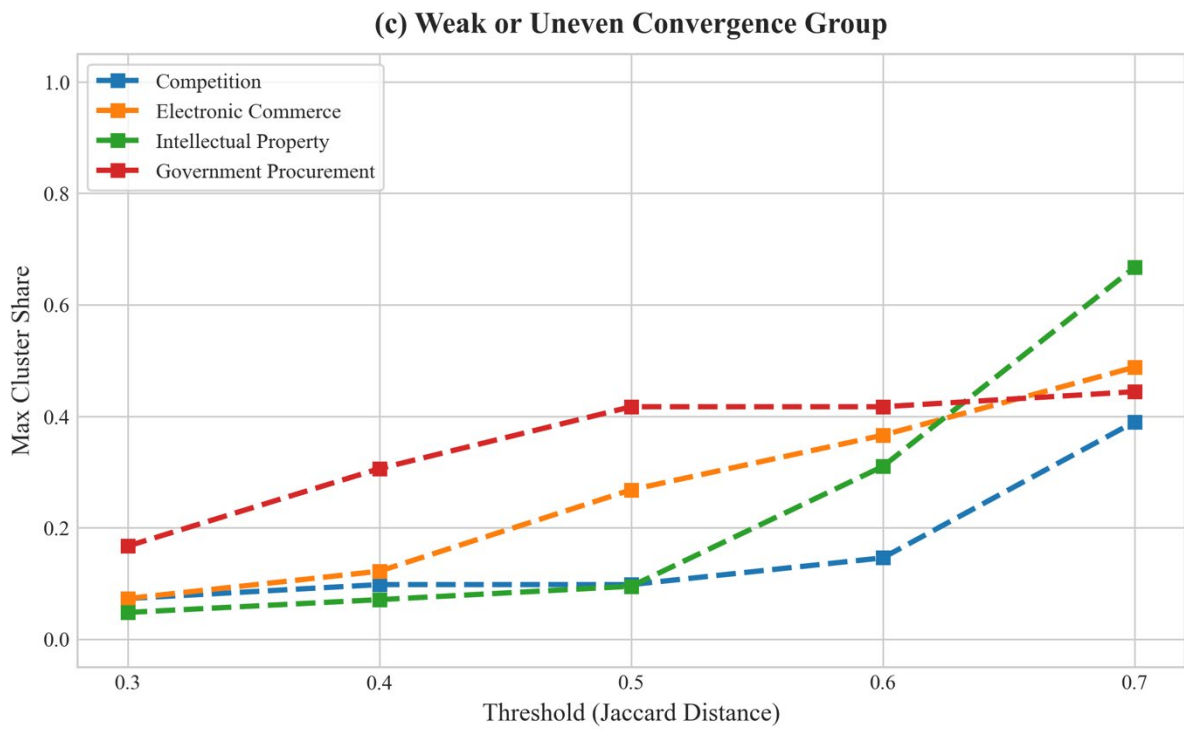
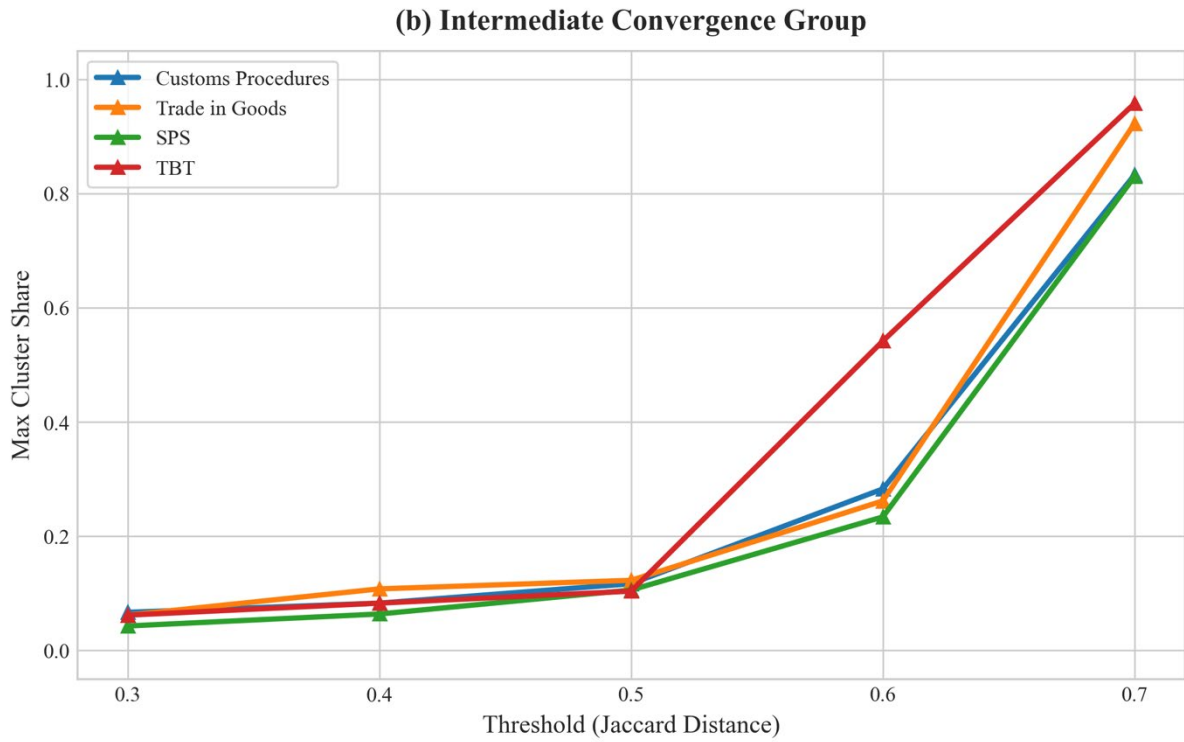
chapters contain some localised textual similarities, while treaty language in this area remains strongly shaped by FTA-specific or party-specific factors.

Electronic commerce also belongs to this weak or uneven convergence group, but for a different reason. As discussed in relation to Figure 4, electronic commerce provisions do not appear to be entirely idiosyncratic. The dendrogram indicates identifiable subclusters corresponding to successive drafting models, including earlier paperless trading provisions, more developed electronic commerce chapters, and more recent digital trade or digital economy frameworks. However, Figure 5 shows that these subclusters do not rapidly combine into a single broad cluster. Even as the threshold is relaxed, the largest cluster share remains relatively limited compared with the early and intermediate convergence groups. This suggests that electronic commerce is characterised by the coexistence of several distinct generations or models of digital rulemaking, rather than by broad textual convergence.

Intellectual property occupies a somewhat different position within the third group. Its largest cluster share remains low at stricter thresholds, but rises more clearly at higher thresholds. This may indicate that a certain degree of broader convergence is emerging in intellectual property chapters, even though it is weaker and slower than in the first and second groups. In this sense, intellectual property appears to be fragmented at the level of detailed wording, while sharing some broader textual or regulatory elements at a more relaxed level of comparison.

Figure 5: Changes in the Largest Cluster Share across Jaccard Distance Thresholds





Source: Compiled and calculated by the author.

Taken together, Figures 3 to 5 provide a more nuanced picture of textual convergence in ASEAN's external FTAs. Figure 3 shows differences in the overall distribution of pairwise similarity across chapters. Figure 4 illustrates how such similarities are organised into specific clusters of FTAs. Figure 5 then shows how quickly these clusters expand as the distance threshold is increased.

The results suggest that textual convergence is not uniform across policy areas. Some fields, such as trade in services, investment, rules of origin, and trade remedies, display relatively high median similarity and early cluster formation. Other fields, such as customs procedures, trade in goods, SPS, and TBT, show intermediate convergence: detailed wording varies, but broader textual and institutional structures are shared by many FTAs. Finally, competition, electronic commerce, intellectual property, and government procurement remain more fragmented or uneven. In particular, electronic commerce illustrates how a field can contain identifiable subclusters without converging into a single broad textual cluster.

These findings also suggest that the architecture of ASEAN's external FTAs is layered rather than homogeneous. Some chapters are shaped by widely diffused legal templates; others are structured by multilateral disciplines, regional treaty practice, or shared institutional frameworks; still others remain dependent on the specific negotiating parties, their domestic legal and policy preferences, or the emergence of new regulatory models. Textual similarity in ASEAN's external FTAs should therefore be understood not as a single process of convergence, but as the coexistence of multiple patterns of standardisation, localisation, and fragmentation across different areas of trade governance.

5. Party-Level Similarity Patterns

5.1 Internal and External Similarity Measures

The previous section examined how textual similarity varies across policy areas and how FTAs cluster within each chapter. This section shifts the focus from chapters to parties. It examines whether particular countries or regions tend to maintain internally consistent treaty language across their FTAs, and whether their agreements are textually close to, or distant from, the broader body of treaty language in the dataset. To do so, it introduces two party-level measures: internal similarity and external similarity.⁹ For the purpose of this party-level analysis, ASEAN+1 agreements, RCEP, and CPTPP

⁹ The idea of examining consistency within a party's treaty network is related to Alschner and Skougarevskiy's analysis of consistency and innovation in national networks of international investment agreements (IIAs) (Alschner and Skougarevskiy, 2016b). Their analysis also examines textual links across treaty networks to explore treaty design diffusion. The internal and external similarity measures used here are conceptually related to these strands of analysis, but are designed as party-level indicators for the present dataset of ASEAN's external FTAs.

are treated as separate plurilateral or regional agreements and are not attributed to individual ASEAN Member States when calculating party-level similarity scores.¹⁰

Internal similarity refers to the average textual similarity amongst FTAs in which a given party participates. This measure is calculated for parties that appear in at least two FTAs within the dataset. For instance, if a party is involved in three FTAs, internal similarity is calculated as the average of the pairwise similarities amongst those three FTAs. This measure captures the degree of consistency in the treaty texts associated with that party. A high internal similarity score suggests that the party tends to maintain a stable treaty template across multiple negotiations, whether that template is distinctive to the party itself or reflects a widely diffused model that the party regularly adopts.

External similarity, by contrast, refers to the average textual similarity between FTAs involving a given party and those in which the party does not participate. This measure captures the extent to which a party's FTAs are textually close to the broader body of treaty language covered by this study. A high external similarity score means that, on average, the party's FTAs are similar to FTAs concluded by other parties in the dataset. In this sense, external similarity may indicate proximity to commonly used or widely diffused treaty language within the dataset. However, a high external similarity score does not necessarily imply active rule-setting power, institutional centrality, or convergence toward a single standard model. It may also result from the party's adoption of treaty language developed by other actors, such as widely diffused regional or international models.

Combining these two measures allows us to map each party's textual position within ASEAN's external FTA network. Rather than creating strict categories, this approach highlights broad tendencies, showing whether a party's FTAs are internally coherent, textually distinctive from the broader treaty space, broadly aligned with commonly used treaty language, or highly variable across negotiations.¹¹

5.2 Mapping Party-Level Similarity

Figure 6 plots each party according to its average internal similarity and average external similarity across the 12 chapters analysed in this paper. The horizontal axis reports external similarity, while the vertical axis reports internal similarity. The dashed

¹⁰ This coding choice is made to avoid attributing ASEAN-wide or plurilateral treaty language to individual ASEAN Member States. ASEAN+1 agreements, RCEP, and CPTPP are therefore not counted as FTAs of individual ASEAN Member States when calculating their internal and external similarity scores, although they remain part of the agreement-level and chapter-level analyses.

¹¹ It should be noted, however, that because the dataset is limited to ASEAN's external FTA network, these measures reflect a party's treaty practice only within this specific FTA space. For instance, a high internal similarity score for an external partner such as Japan or the UAE indicates consistency in its FTAs with ASEAN and ASEAN Member States, but does not necessarily imply that the same template is maintained in its FTAs with other countries or regions. Likewise, external similarity measures relative proximity within this dataset, rather than a party's centrality in the wider universe of FTAs.

lines indicate the mean values of the plotted parties and are used only as visual reference points, rather than as strict classification thresholds. While all parties with at least two FTAs in the dataset are plotted, care should be taken when interpreting the positions of parties with a limited number of FTAs.¹²

The purpose of the figure is therefore not to assign parties to fixed types, but to provide a descriptive map of their relative textual positions. By examining internal and external similarity together, the figure shows the extent to which each party's FTAs are internally coherent and how closely they are connected to the broader body of treaty language in the dataset.

Amongst the partner countries and regions, Japan, the EU, EFTA, and the EAEU are located in the upper-left area of the figure, combining above-average internal similarity with below-average external similarity. This suggests that their FTAs are relatively coherent internally, while retaining textual features that are less widely shared across the broader set of FTAs. This pattern is most pronounced for Japan and the EU, while EFTA and the EAEU display the same general tendency to a more moderate degree.

By contrast, partner countries and regions located in the upper-right area, such as the UAE, Chile, Sri Lanka, and India, show relatively high internal similarity as well as relatively high external similarity. Their FTAs are therefore relatively coherent internally, but also close to the broader body of treaty language in the dataset. This pattern may reflect the use of widely diffused treaty language or regionally and internationally shared drafting models, rather than the maintenance of highly distinctive treaty templates of their own.

Other partner countries and regions, including Türkiye, the Republic of Korea, China, Australia, and New Zealand, are located closer to the lower or central part of the figure, with relatively lower internal similarity. This suggests that the FTAs concluded by these parties in the dataset vary more substantially from one another. In other words, their treaty practice appears less internally uniform, possibly reflecting differences in negotiating partners, time periods, and FTA frameworks.

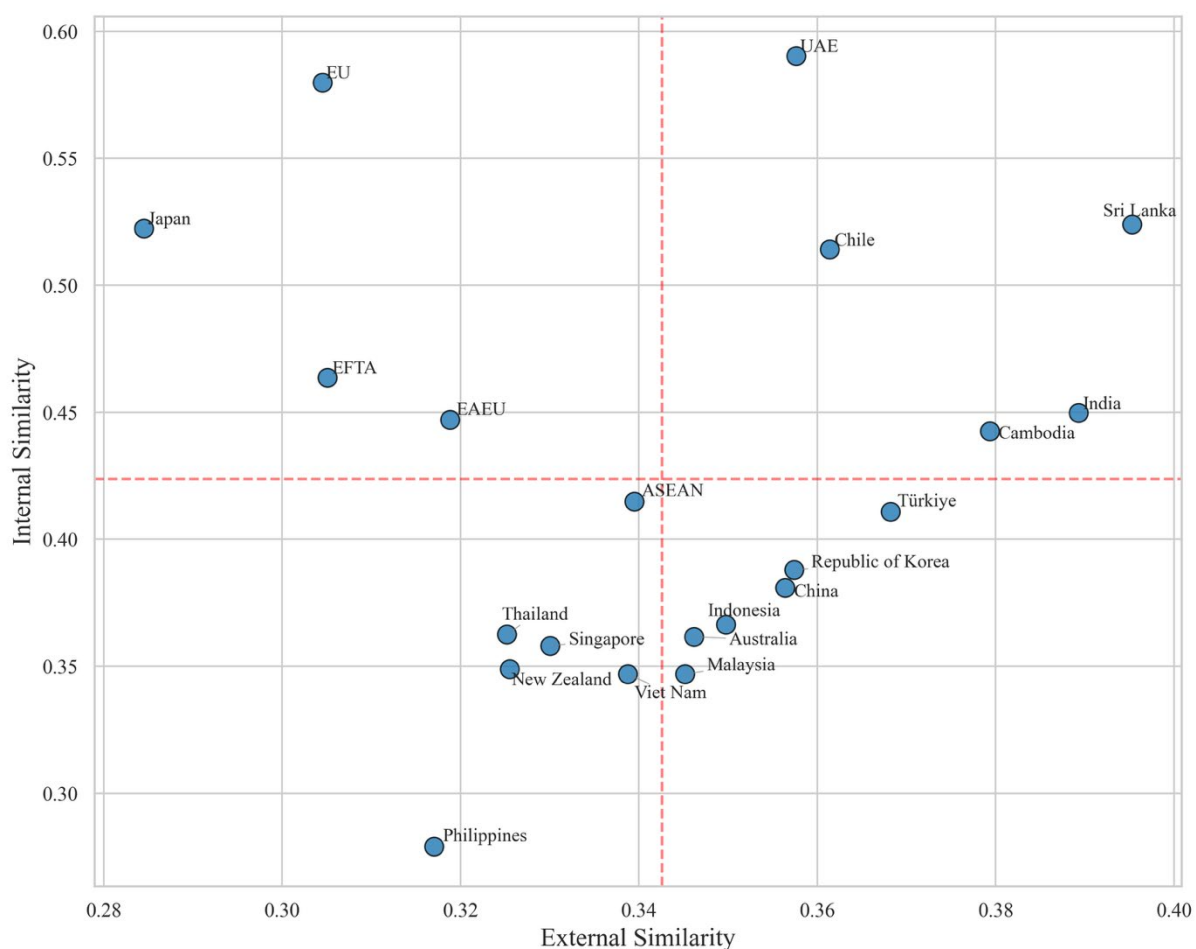
A similar pattern can be observed amongst many ASEAN Member States, which are also located in the lower or central areas of the figure. This is consistent with the fact that ASEAN Member States have concluded bilateral FTAs with diverse partners, at different points in time, and under varying negotiating contexts. At the same time, Malaysia and Indonesia have external similarity scores above the sample mean, indicating that their FTAs are relatively close to the broader FTA space, even though their internal similarity remains modest.

The Philippines and Cambodia stand somewhat apart from the main concentration of ASEAN Member States, although both have only three bilateral FTAs in the dataset. Their positions are nevertheless contrasting. The Philippines shows particularly low internal

¹² Specifically, the EU, Sri Lanka, and Türkiye are involved in only two FTAs within the dataset.

similarity, which may reflect the diversity of its FTAs with Japan, EFTA, and the Republic of Korea. These agreements were signed over a wide time span, from the 2006 agreement with Japan to the 2023 agreement with the Republic of Korea, and involve partners with different treaty practices. Cambodia, by contrast, is located in the upper-right area of the figure. Its three FTAs, with China, the Republic of Korea, and the UAE, were signed between 2020 and 2023. Its relatively high internal and external similarity may therefore partly reflect the concentration of these agreements within a recent period, when treaty language and drafting models had become more widely diffused across regional and international FTAs.

Figure 6: Internal and External Similarity by Party



Note: The figure includes all parties that are involved in at least two FTAs in the dataset. ASEAN+1 FTAs, RCEP, and CPTPP are not attributed to individual ASEAN Member States when calculating party-level similarity scores. Because the number of FTAs differs substantially by party, results for parties with only two or three FTAs should be interpreted with particular caution.

Source: Compiled and calculated by the author.

ASEAN, treated here as a collective contracting party, is located near the centre of the scatter plot. Its internal and external similarity scores are both close to the mean values of the plotted parties. This intermediate position should not be overinterpreted as directly measuring ASEAN's institutional role, but it is consistent with the character of

ASEAN+1 FTAs as collective agreements formed between ASEAN's diverse member states and external partners. ASEAN's FTAs do not combine internal coherence and external distinctiveness to the same extent as those of Japan and the EU, or, to a more moderate degree, EFTA and the EAEU, but they are also not as textually dispersed as some FTAs concluded by individual ASEAN Member States. Rather, ASEAN occupies an intermediate position in the FTA space, consistent with both the common structure of ASEAN+1 FTAs and their proximity to broader regional treaty practice.

Overall, the figure should be read not as a rigid classification of parties, but as a descriptive map of their relative textual positions. It shows that parties differ not only in the extent to which their own FTAs resemble one another, but also in the degree to which those FTAs are close to, or distant from, the wider body of treaty language found in ASEAN's external FTAs.

6. Conclusion

This paper has quantitatively compared the textual similarity of 65 FTAs in ASEAN's external FTA network across 12 chapters. Specifically, it calculated character 5-gram-based Jaccard similarity scores and examined the resulting patterns through several complementary approaches: similarity distributions by policy area, hierarchical clustering, the largest cluster share, and party-level internal and external similarity.

First, the degree of textual convergence in ASEAN's external FTAs differs substantially across policy areas. Trade in services, investment, rules of origin, and trade remedies show relatively high similarity amongst FTA pairs and tend to form large clusters at relatively strict distance thresholds. In these areas, the observed patterns are consistent with a relatively strong influence of earlier treaty models and standardised drafting structures on subsequent FTAs. Such shared models may derive from existing multilateral rules, such as the WTO agreements, from bilateral investment treaty practice, or from previously established FTA templates.

Second, trade in goods, customs procedures, SPS, and TBT display a more intermediate pattern of convergence. At strict textual thresholds, these areas still show a certain degree of variation. However, once the distance threshold is relaxed, the share of FTAs belonging to the largest cluster rises relatively quickly. This suggests that, although individual FTAs retain differences in drafting style and textual expression, their textual formulations appear to be built around broadly shared institutional frameworks and basic chapter structures. These areas may therefore be understood not as fully standardised fields, but as fields of moderate convergence, in which country- or partner-specific adjustments are made on the basis of broadly shared institutional foundations.

Third, competition, electronic commerce, intellectual property, and government procurement show lower overall similarity and slower or more limited formation of large clusters. In these areas, domestic regulatory systems, policy objectives,

negotiation timing, and partner-specific preferences may be more directly reflected in treaty drafting, resulting in greater textual diversity. At the same time, these fields should not be treated as a single homogeneous group. In electronic commerce and intellectual property, the largest cluster share rises when the threshold is relaxed, suggesting that some degree of convergence may be emerging, particularly amongst more recent FTAs. Government procurement, by contrast, shows a somewhat different pattern: although some FTA pairs exhibit high similarity, these localised similarities do not necessarily develop into broader convergence across the field.

Fourth, the cluster analysis identified several distinctive groups of FTAs across policy areas. FTAs involving Japan and the UAE frequently formed compact partner-centred clusters, a pattern consistent with the use of recognisable treaty templates across multiple agreements. Clusters centred on individual ASEAN Member States were also observed, most recurrently for Singapore, with more limited examples for Malaysia, although their composition varied across chapters. A separate and more limited ASEAN-wide pattern was found in trade in goods, trade in services, and investment, where several ASEAN+1 FTAs formed relatively coherent initial clusters. RCEP and CPTPP also showed textual proximity to certain other FTAs in a number of policy areas. These patterns are consistent with the circulation of treaty language through bilateral, regional, and mega-regional agreement networks, although textual similarity alone cannot establish the existence or direction of such influence.

Fifth, the analysis of internal and external similarity made it possible to characterise treaty parties along two dimensions: the consistency of their own treaty practice and their proximity to the broader treaty space. The results show that some partner countries and regions, including Japan, the EU, EFTA, and the EAEU, combine relatively high internal similarity with relatively low external similarity, suggesting that their FTAs are internally coherent while retaining distinctive textual features. Other partners, including the UAE, Chile, Sri Lanka, and India, also show relatively high internal similarity, but their FTAs are closer to the broader body of treaty language in the dataset. By contrast, a number of ASEAN Member States and other partner countries and regions are located closer to the lower or central part of the figure, reflecting greater variation across FTAs concluded with different partners, at different times, and in different negotiating contexts. ASEAN, treated as a collective contracting party, occupies an intermediate position, with both its internal and external similarity scores close to the sample means. This pattern is consistent with the collective character of ASEAN+1 FTAs, which link ASEAN's diverse member states with external partners.

Taken together, these findings show that convergence in ASEAN's external FTAs cannot be adequately captured by overall or average similarity scores alone. The degree of convergence, the pace at which large clusters emerge, and the breadth of textual similarity vary substantially across policy areas and treaty parties. Moreover, even where high similarity is observed, it is necessary to distinguish between broad field-

wide convergence and localised similarity generated by the treaty templates of particular parties or by specific negotiating contexts.

The analysis also has several limitations. First, character 5-gram-based Jaccard similarity captures formal textual overlap; it does not directly assess legal meaning, substantive equivalence, or the strength of legal obligations. Similar wording may produce different legal effects depending on exceptions, schedules of reservations, scope of application, institutional arrangements, or dispute settlement coverage. Conversely, provisions with low textual similarity may still perform similar legal functions. Nor can textual similarity alone establish the direction of influence amongst FTAs. The results should therefore be understood as an initial textual map, not as a substitute for detailed legal interpretation.

Second, the dataset is limited to ASEAN's external FTAs. The patterns observed in this paper may not necessarily apply to FTAs in other regions or to the broader universe of global trade agreements. Accordingly, the party-level measures describe a party's treaty practice only within this specific FTA network and should not be interpreted as measures of its overall treaty practice or its centrality in the wider universe of FTAs. In addition, party-level measures should be interpreted cautiously where the number of FTAs associated with a particular party is small, since internal and external similarity scores may be strongly affected by one or two agreements.

Several avenues for future research follow from these limitations. One potential avenue is to move beyond chapter-level similarity and conduct detailed provision-level analyses. Structuring the data at the provision level would allow future research to examine textual convergence in much greater detail, including whether specific obligations, exceptions, or institutional provisions diffuse across agreements in different ways.

Another important step would be to expand the scope of the dataset, both in terms of policy areas and the range of FTAs. Regarding policy areas, future research could expand beyond the 12 chapters analysed in this study to include emerging fields such as sustainability and small and medium-sized enterprises (SMEs), which were excluded here because of their small sample sizes. As for the scope of agreements, future research could include not only ASEAN's external FTAs, but also FTAs concluded by ASEAN's external partners with non-ASEAN countries or regions. This would make it possible to distinguish more clearly between features specific to ASEAN's external FTA network and those reflecting broader patterns in global FTA text convergence. Additionally, comparing ASEAN's external FTAs with intra-ASEAN agreements, such as ATIGA, ATISA, and ACIA, would shed light on the interaction between ASEAN's external FTAs and ASEAN's own internal treaty models.

This paper has shown that ASEAN's external FTAs are not converging in a single, uniform direction. Rather, they display multiple patterns of convergence and divergence depending on the policy area, treaty party, and agreement framework. This finding suggests that ASEAN's external FTA network cannot be fully understood through broad

labels such as ‘deep trade agreements’ or ‘high-standard trade agreements.’ A more nuanced understanding requires attention to the textual structure of individual policy areas and to the treaty practices of particular parties. Text-as-data analysis offers a useful supplementary approach for mapping this complex FTA network and can contribute to future research on FTA design, regional economic integration, and ASEAN’s evolving external economic relations.

References

- Allee, T. and M.Elsig (2019), ‘Are the Contents of International Treaties Copied and Pasted? Evidence from Preferential Trade Agreements’, *International Studies Quarterly*, 63(3), pp.603–13.
- Allee, T. and A.Lugg (2016), ‘Who Wrote the Rules for the Trans-Pacific Partnership?’, *Research & Politics*, 3(3).
- Allee, T., M. Elsig, and A.Lugg (2017a), ‘Is the European Union Trade Deal with Canada New or Recycled? A Text-as-Data Approach’, *Global Policy*, 8(2), pp.246–52.
- Allee, T., M. Elsig, and A.Lugg (2017b), ‘The Ties between the World Trade Organization and Preferential Trade Agreements: A Textual Analysis’, *Journal of International Economic Law*, 20(2), pp.333–63.
- Alschner, W. and D.Skougarevskiy (2016a), ‘The New Gold Standard? Empirically Situating the TPP in the Investment Treaty Universe’, *Journal of World Investment & Trade*, 17(3), pp.339–73.
- Alschner, W. and D.Skougarevskiy (2016b), ‘Mapping the Universe of International Investment Agreements’, *Journal of International Economic Law*, 19(3), pp.561–88.
- Alschner, W., J.Seiermann, and D.Skougarevskiy (2017), ‘Text-as-Data Analysis of Preferential Trade Agreements: Mapping the PTA Landscape’, *UNCTAD Research Paper* No. 5, UNCTAD/SER.RP/2017/5/Rev.1, United Nations Conference on Trade and Development (UNCTAD), Geneva.
- ASEAN Geoeconomics Task Force (2025), ASEAN Geoeconomics Report 2025: This Time It Is Different: ASEAN’s Agency in Shaping the New Global Agenda. October 2025.
- Asian Development Bank (2022), *The Regional Comprehensive Economic Partnership Agreement: A New Paradigm in Asian Regional Cooperation*. Tokyo: Asian Development Bank.
- Baccini, L., A.Dür, and Y.Z.Haftel (2015), ‘Imitation and Innovation in International Governance: The Diffusion of Trade Agreement Design’, in A. Dür and M. Elsig

- (eds), *Trade Cooperation: The Purpose, Design and Effects of Preferential Trade Agreements*, Cambridge: Cambridge University Press, pp.167–94.
- Frank, N. (2023), 'Power by Proxy: Explaining Innovation and Imitation in the RCEP', *Global Governance: A Review of Multilateralism and International Organizations*, 29(3), pp.322–45.
- Kawai, M. and G.Wignaraja (2011), 'Asian FTAs: Trends, Prospects, and Challenges', *Journal of Asian Economics*, 22(1), pp.1–22.
- Kleimann, D. (2013), 'Beyond Market Success? The Anatomy of ASEAN's Preferential Trade Agreements', EUI Working Papers No. LAW 2013/01, European University Institute.
- Lee, J. and K.G.Tan (2025), 'From Innovation to Diffusion: Cross-Regional Influence of the US-Singapore FTA', ACI Research Paper No. 09-2025, Asia Competitiveness Institute Research Paper Series, Lee Kuan Yew School of Public Policy, National University of Singapore.
- Malingrey, L. and Y.Duval (2022), 'Mainstreaming Sustainable Development in Regional Trade Agreements: Comparative Analysis and Way Forward for RCEP', *ARTNeT Working Paper Series* No. 213, Asia-Pacific Research and Training Network on Trade (ARTNeT), Bangkok.
- Mattoo, A., N.Rocha, and M.Ruta (eds.) (2020), *Handbook of Deep Trade Agreements*. Washington, DC: World Bank.
- Morin, J.-F., D.Blümer, C.Brandi, and A.Berger (2019), 'Kick-Starting Diffusion: Explaining the Varying Frequency of Preferential Trade Agreements' Environmental Provisions by Their Initial Conditions', *The World Economy*, 42, pp.2602–28.
- Ravenhill, J. (2016), 'The Political Economy of an "Asian" Mega-FTA: The Regional Comprehensive Economic Partnership', *Asian Survey*, 56(6), pp.1077–100.
- Rillo, A.D., A.M.R.D.Robeniol, and S.M.Buban (2022), 'The Story of RCEP: History, Negotiations, Structure, and Future Directions', in F. Kimura, S.Thangavelu, and D.Narjoko (eds.), *Regional Comprehensive Economic Partnership: Implications, Challenges, and Future Growth of East Asia and ASEAN*, pp.83–110. Jakarta: ERIA. https://www.eria.org/uploads/media/Books/2022-RCEP-Book1/6_Ch.2-History-of-RCEP.pdf
- Serafica, R.B. and I.M.Ramli (2022), 'RCEP Services Liberalisation: Key Features and Implications', in F.Kimura, S.Thangavelu, and D.Narjoko (eds.), *Regional Comprehensive Economic Partnership: Implications, Challenges, and Future Growth of East Asia and ASEAN*, pp.83–109. Jakarta: ERIA. https://www.eria.org/uploads/media/Books/2022-RCEP-Book1/8_Ch.4-RCEP-services-liberalisation.pdf

Appendix

Table A: List of Signatories to the FTAs Covered in this Study

Category	ASEAN Party / Framework	Number of FTAs	Partners (Signature Year)
Bilateral FTAs (Total: 57)	Singapore	20	New Zealand (2000), Japan (2002), EFTA (2002), US (2003), Jordan (2004), Republic of Korea (2005), India (2005), Panama (2006), China (2008), Peru (2008), GCC (2008), Costa Rica (2010), Chinese Taipei (2013), Türkiye (2015), Australia (2017), EU (2018), Sri Lanka (2018), EAEU (2019), Pacific Alliance (2022), Mercosur (2023)
	Malaysia	9	Japan (2005), Pakistan (2007), New Zealand (2009), Chile (2010), India (2011), Australia (2012), Türkiye (2014), UAE (2025), EFTA (2025)
	Indonesia	8	Japan (2007), Chile (2017), EFTA (2018), Australia (2019), Republic of Korea (2020), UAE (2022), Canada (2025), EAEU (2025)
	Viet Nam	7	Japan (2008), Chile (2011), Republic of Korea (2015), EAEU (2015), EU (2019), Israel (2023), UAE (2023)
	Thailand	6	Australia (2004), New Zealand (2005), Japan (2007), Chile (2013), Sri Lanka (2024), EFTA (2025)
	Cambodia	3	China (2020), Republic of Korea (2021), UAE (2023)
	Philippines	3	Japan (2006), EFTA (2016), Korea (2023)
	Brunei	1	Japan (2007)
Mega-Regional FTAs (Total: 2)	RCEP	1	15 nations (incl. 10 ASEAN members) (2020)
	CPTPP	1	11 nations (incl. 4 ASEAN members) (2018)
ASEAN+1 FTAs (Total: 6)	ASEAN	6	China (2004), Republic of Korea (2006), Hong Kong (2007), Japan (2008), Australia/New Zealand (2009), India (2009)

Note: Signature years refer to the original signature year of the FTA covered in this study, rather than the year of subsequent upgrades, amendments, or protocols. For the ASEAN+1 FTAs, the signature years refer, where applicable, to the relevant trade-in-goods agreements or comprehensive FTA texts covered in this study, rather than earlier framework agreements.

Source: Compiled by the author based on the Japan External Trade Organization (JETRO), *World FTA Database* (in Japanese).

Table B: Top 50 FTA Pairs by Cross-Sectoral Mean Jaccard Similarity

FTA 1	FTA 2	Mean Jaccard Similarity	Number of Shared Chapters
Indonesia–Japan	Brunei–Japan	0.781	5
Malaysia–Japan	Brunei–Japan	0.752	5
Brunei–Japan	Viet Nam–Japan	0.729	4
Malaysia–UAE	Viet Nam–UAE	0.716	11
Thailand–Japan	Brunei–Japan	0.706	5
CPTPP	Singapore–Pacific Alliance	0.706	10
Brunei–Japan	Philippines–Japan	0.7	5
Viet Nam–Japan	ASEAN–Japan	0.697	5
Malaysia–UAE	Cambodia–UAE	0.69	10
Singapore–Türkiye	Singapore–Sri Lanka	0.661	12
Viet Nam–UAE	Cambodia–UAE	0.661	10
Malaysia–Japan	Thailand–Japan	0.649	7
Malaysia–Japan	Indonesia–Japan	0.648	7
Malaysia–Japan	Philippines–Japan	0.647	7
ASEAN–China	ASEAN–Republic of Korea	0.647	4
ASEAN–India	ASEAN–Republic of Korea	0.633	4
Viet Nam–Republic of Korea	Indonesia–Republic of Korea	0.633	6
Philippines–Republic of Korea	Cambodia–Republic of Korea	0.632	4
RCEP	Philippines–Republic of Korea	0.629	5
ASEAN–Australia, New Zealand	RCEP	0.615	12
ASEAN–China	ASEAN–India	0.613	4
Indonesia–Republic of Korea	Cambodia–Republic of Korea	0.611	4
Indonesia–Japan	Viet Nam–Japan	0.61	6
Singapore–Peru	Singapore–Costa Rica	0.605	11
RCEP	Cambodia–Republic of Korea	0.598	4
Malaysia–Chile	Viet Nam–Chile	0.597	6
Singapore–Panama	Singapore–Costa Rica	0.595	10
Thailand–Japan	Philippines–Japan	0.595	9
Philippines–Japan	Viet Nam–Japan	0.589	6
Brunei–Japan	ASEAN–Japan	0.583	4
Thailand–Japan	ASEAN–Japan	0.582	4
ASEAN–Australia, New Zealand	Philippines–Republic of Korea	0.58	5
Singapore–EU	Viet Nam–EU	0.58	12
Thailand–Japan	Indonesia–Japan	0.579	9
Indonesia–Japan	ASEAN–Japan	0.579	4
Singapore–Japan	Brunei–Japan	0.577	5
Thailand–Australia	Thailand–New Zealand	0.573	12
Thailand–Japan	Viet Nam–Japan	0.573	6
CPTPP	Indonesia–Canada	0.573	12
Viet Nam–Chile	Thailand–Chile	0.57	6
ASEAN–China	RCEP	0.57	9
Philippines–Japan	ASEAN–Japan	0.565	4
Singapore–Pacific Alliance	Indonesia–Canada	0.562	10

Viet Nam–Republic of Korea	Cambodia–Republic of Korea	0.56	4
Indonesia–Republic of Korea	Philippines–Republic of Korea	0.559	4
ASEAN–Australia, New Zealand	Indonesia–Australia	0.556	9
Indonesia–EFTA	Thailand–EFTA	0.555	8
Singapore–Panama	Singapore–GCC	0.555	6
Malaysia–Japan	Viet Nam–Japan	0.551	8
Malaysia–Chile	Thailand–Chile	0.55	6

Note: Mean Jaccard Similarity is calculated as the average of the pairwise similarity scores across the shared chapters between the two FTAs. FTAs are consistently labeled in the format of 'ASEAN party–partner country or region.' The total number of FTAs varies across parties (with Japan having the maximum of seven). Consequently, parties with more FTAs are naturally more likely to appear frequently in this top-50 ranking due to a larger number of potential pairwise combinations.

Source: Compiled and calculated by the author.