Socio-Technical Configuration Analysis of Space Objects for Enhanced Space Domain Awareness

Tiffany M. Phan
The University of Texas at Austin
Moriba K. Jah
The University of Texas at Austin

ABSTRACT

This work applies natural language processing and network analysis to open-source discourse on China's satellite operations (2005–2025). From 500 documents, over 1,000 actor–relation–concept triplets were extracted and organized into time-sliced socio-technical graphs, revealing shifts from transparency and technological leadership toward deterrence and ambiguity. The NLP-derived patterns aligned with an independent historical periodization, indicating the method's validity. This approach shows how structured analysis of soft data can enhance real-time monitoring, threat assessment, and policy insight in space domain awareness.

1. INTRODUCTION

A critical task of space domain awareness (SDA) is the detection and prediction of threats to anthropogenic space objects (ASOs) in Earth's orbit. Computer-based SDA is primarily approached from a hard data standpoint; the position and trajectory of an ASO are the primary data of interest for simulation and analysis. However, space operators need to know not only where objects are at any time, but also how objects arrived in orbit, who controls them, what their capabilities might be, and what intentions their operators may hold [22]. This contextual information, which cannot be gathered from physical sensors like telescopes or radar, is vital for evaluating the intent, opportunity, and capability of actors to interfere with or harm space assets.

Despite its importance, soft data has historically been underutilized in SDA due to its unstructured nature, the complexity of extraction processes, and the substantial human effort required to analyze it. However, recent advances in natural language processing (NLP) such as the development of attention mechanisms, transformer architectures, and large language models (LLMs) offer scalable and automated methods for extracting and synthesizing soft data from diverse textual sources.

One promising application of soft data is the construction and analysis of socio-technical configurations of ASOs. These configurations map the relationships among actors, technologies, and institutions by analyzing event data, technical discourse, and policy narratives. The goal is not only to identify physical networks of collaboration, but to understand the evolving worldviews, practices, and institutional alignments that influence space activities. Such analysis provides insight into how new socio-technical configurations emerge, diffuse, and scale, ultimately enhancing situational awareness in the orbital domain by contextualizing the use, purpose, and ownership of space-based assets.

This research adopts a substantive approach to socio-technical configuration, focusing on observable activities and institutional changes enacted by various actors. Given the volume and complexity of relevant textual data, traditional manual methods are infeasible at scale. This work therefore proposes a scalable NLP framework to automate the extraction of socio-technical relationship information from unstructured text, enabling timely and comprehensive analysis in support of SDA.

2. BACKGROUND AND LITERATURE REVIEW

2.1 Space Domain Awareness and the Role of Contextual Information

Space Domain Awareness (SDA) is the discipline concerned with understanding the space environment, predicting potential threats, and supporting safe and sustainable use of Earth's orbit. With the proliferation of anthropogenic space objects (ASOs)—including operational satellites, defunct spacecraft, and debris—the orbital domain has become increasingly congested and contested. Thousands of satellites serve vital functions such as global positioning, climate monitoring, intelligence gathering, and telecommunications, while their presence increases the likelihood of catastrophic collisions. Such events, like the 2009 Iridium-Cosmos collision, exemplify the risk of cascading debris effects known as the Kessler Syndrome, which hypothesizes that each collision generates further debris that threatens other ASOs [16].

To mitigate these risks, SDA traditionally relies on hard data obtained through physical means, such as telescopes, radars, and onboard sensors. These methods provide valuable information on the kinematic state of objects, including their position, velocity, and orbital characteristics. While these data are essential, they offer little insight into the context of objects: who owns them, how they were launched, what their intended purpose is, and what strategic or operational intentions their operators may have [22]. This kind of soft data (contextual information derived from policies, public communications, or institutional affiliations) is essential for inferring actor capability, intent, and opportunity. It is this information that often underpins assessments of geopolitical risk, deterrence, and conflict prediction in space.

However, soft data has historically been underexploited in SDA due to its reliance on unstructured text and human analysts. Incomplete or outdated satellite registries, such as the United Nations Register of Objects Launched into Outer Space, further exacerbate the challenge [13]. As the launch cadence accelerates and actors diversify, scalable computational tools are urgently needed to extract and analyze contextual data.

2.2 Natural Language Processing for Soft Data Extraction

Recent advances in natural language processing (NLP) have made it increasingly feasible to extract structured knowledge from unstructured text. In particular, transformer-based models such as BERT, RoBERTa, and their question-answering (QA) variants have demonstrated strong performance in tasks including named entity recognition (NER), relation extraction (RE), and event detection.

Despite their wide applicability, few efforts have addressed the challenge of applying these techniques to socio-technical analysis in the orbital domain. The complexity of this domain lies not only in the diversity of actors (governmental, commercial, and civil), but also in the layered relationships that span technical, legal, and geopolitical dimensions.

Let $T = \{t_1, t_2, \dots, t_N\}$ denote a corpus of unstructured text documents (e.g., news articles, reports, and press releases). The goal is to extract entities E and semantic relations R from this corpus to construct a knowledge graph $G = (E, R)$, representing socio-technical configurations of actors and objects in the space domain.

We adopt a large language model (LLM) or question-answering (QA) approach for this task. Formally, let $\Phi_{\text{LLM}} : (q, c) \mapsto a$ be a model that maps a query q and a context window $c \subseteq T$ to an answer span a, where $a \in \mathbb{S}$, the space of text spans. This formulation supports prompt-based extraction where prompts are structured to elicit actor, action, and object triplets from context.

For each document t_i, the system iteratively applies a set of extraction prompts $Q = \{q_1, q_2, \dots, q_k\}$ to obtain tuples (e_i, r_{ij}, e_j). These are accumulated to build the evolving knowledge graph G. The model parameters θ are optimized to maximize:

$$\max_{\theta} \sum_{i=1}^{N} \sum_{q \in Q} \log P(a_{iq} \mid q, t_i; \theta),$$

where a_{iq} is the answer span produced for query q on document t_i.

LLMs such as BERT [7], RoBERTa [18], or domain-adapted variants [10] use contextual embedding functions $\phi : T \to \mathbb{R}^d$ and multi-head self-attention [24] to capture co-reference and long-distance dependencies, which are critical in extracting latent socio-technical narratives. These embeddings enable the model to resolve references to the same actor across documents and to identify implicit causal or organizational relations.

This approach is particularly well-suited to modeling socio-technical systems, where actor-technology-value relationships are not always explicit and may be distributed across multiple sources and contexts.

2.3 Socio-Technical Configuration Analysis (STCA)

To interpret soft data meaningfully, we draw on theoretical insights from *Socio-Technical Configuration Analysis (STCA)* [19]. STCA is a framework from transition studies and innovation systems research that emphasizes the interdependencies among actors, institutions, technologies, and cultural practices. It enables the identification and tracking of dominant configurations, or "regimes," that shape the evolution of technological fields [11].

STCA provides a systematic, semi-quantitative method for understanding how configurations emerge, stabilize, or dissolve over time and across geographies. Its core strength lies in capturing not just formal institutional alignments, but also the narratives, worldviews, and storylines that actors co-construct and use to legitimize their actions. These storylines function as templates that reinforce specific configurations and may reflect or predict shifts in socio-technical systems [21]. By capturing these narratives, STCA allows for a deeper understanding of how socio-technical systems evolve, often in ways that are not strictly driven by technological innovation alone, but also by shifts in values and norms.

The STCA methodology includes a diverse range of components:

- **Material components**: technologies, infrastructure, and tangible artifacts.
- **Institutional components**: governance structures, norms, values, and organizational practices.
- **Associative relationships**: connections between concepts or actors, identified through their simultaneous appearance in discourse or action.
- **Actor-Congruence Networks**: networks formed by actors sharing ideational similarities, supporting specific narratives or configurations, even without direct collaboration.

These features align well with capabilities afforded by NLP. Statements linking actors to actions, policies, or technologies can be modeled as triplets, while co-occurrence patterns in text provide the basis for constructing actor-congruence networks or concept co-appearance graphs.

For instance, multiple actors referencing "space debris mitigation" within distinct but related statements could signal the coalescence of a proto-regime around orbital sustainability. Clustering such associations reveals dominant storylines, which serve as proxies for broader socio-technical configurations. By comparing these clusters over time, STCA enables comparative studies of technological change, policy evolution, and regime transformation, making it ideal for applications in the emerging space economy [2].

Formally, we define a socio-technical configuration C as a multigraph:

$$C = (A, K, L),$$

where:

- A is the set of actors (government agencies, corporations, etc.),
- K is the set of concepts (technologies, norms, rules, etc.),
- $L \subseteq (A \times K) \cup (A \times A) \cup (K \times K)$ are labeled links derived from co-appearances or references in text.

Let $\sigma : T \to 2^L$ be a mapping from text to links using NLP-based parsing. Over the corpus T, we define an actor congruence network $G_A = (A, E_A)$, where:

$$(a_i, a_j) \in E_A \iff \text{Jaccard}(K_{a_i}, K_{a_j}) > \tau,$$

with K_{a_i} the set of concepts linked to actor a_i, and $\tau \in [0, 1]$ a similarity threshold.

Clusters in G_A represent shared narratives or aligned positions, serving as proxies for "proto-regime" structures. These configurations evolve across time, modeled as a sequence $\{C^{(t)}\}_{t=1}^{T}$ where regime stability or transitions are studied via:

$$\Delta(C^{(t)}, C^{(t+1)}) = |L^{(t+1)} \setminus L^{(t)}| + |L^{(t)} \setminus L^{(t+1)}|$$

where:

- $C^{(t)}$ represents the socio-technical configuration at time t, which includes the relationships between actors, technologies, values, and behaviors.
- $L^{(t)}$ is the set of links (or relationships) in the socio-technical configuration at time t, representing the connections between actors, technologies, and values.
- $L^{(t+1)}$ is the set of links in the configuration at time $t+1$, representing the updated set of relationships after the transition.
- $L^{(t+1)} \setminus L^{(t)}$ counts the new links that have appeared in the configuration at time $t+1$ but were not present at time t.
- $L^{(t)} \setminus L^{(t+1)}$ counts the links that were present in the configuration at time t but no longer appear in the configuration at time $t+1$.
- $\Delta(C^{(t)}, C^{(t+1)})$ is the total number of changes (new and removed links) between the two configurations. It quantifies the shift in the socio-technical regime from time t to $t+1$.

The configuration changes, or the number of added and removed links, reflect shifts in the socio-technical landscape, such as changes in actors' alignments, the introduction of new technologies, or evolving values.

2.4 Narrative Generation Using Generative AI

To make sense of the complex patterns found in socio-technical data, this research includes a narrative generation step that turns structured information into readable summaries. These summaries, or *narratives*, describe how different actors—such as countries or companies—relate to key technologies, values, or behaviors over time.

The input to this process is a network of extracted triplets (e.g., (Russia, promotes, space sustainability)) identified by our NLP system. These triplets are grouped by actor and time period, then organized into themes such as technology development, value alignment, or international cooperation.

To convert this structured data into natural language, we use a generative AI model. This model creates short, coherent text passages that explain how an actor's behavior has changed, what values they emphasize, and how they interact with others. For example, it might generate a paragraph showing how a country has shifted its focus from military surveillance to space debris mitigation.

This automated narrative generation is inspired by research in strategic communication and discourse analysis [6, 12], and it allows analysts to quickly grasp the bigger picture without reading hundreds of individual data points. By combining structured data with generative AI, our framework produces consistent, scalable summaries that support strategic insight in space policy and security.

2.5 Towards an Integrated Framework

Despite the potential of STCA and modern NLP tools, few efforts have integrated them to support scalable SDA. Existing STCA approaches are often labor-intensive, requiring expert coding of text and manual construction of actor networks. Conversely, many NLP pipelines ignore the theoretical underpinnings that make socio-technical interpretation meaningful.

This project bridges that gap by developing a novel, scalable NLP framework that automates entity and relation extraction from space-relevant text and grounds it in STCA theory. The proposed system identifies actors, technologies, concepts, and relationships, and organizes them into socio-technical configurations, which can then be analyzed across time and contexts to support SDA.

This integration enables new forms of insight into the space domain: not only tracking *where* objects are and *what* they are doing, but also *who* is doing it, *why*, and *with what implications*. In doing so, it contributes a novel methodology for the fusion of soft and hard data in service of more robust and context-aware space domain awareness.

3. RESEARCH GOAL

The goal of this research is to build a reproducible, scalable framework for extracting and analyzing socio-technical dynamics in space operations by combining natural language processing (NLP), knowledge graph construction, and narrative analysis. The specific objectives are:

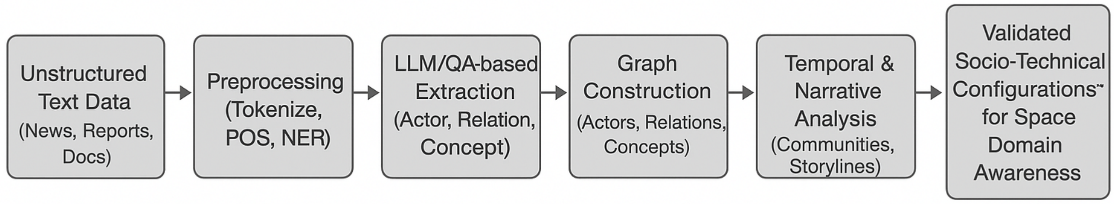

Fig. 1: NLP pipeline integrating LLM-based extraction with socio-technical analysis

1. **Identify and Curate Relevant Source Material**: Collect and structure data related to space operations. This includes government documents, military and technical analyses, commercial space surveillance reports, and journalistic coverage.
2. **Retrieve and Preprocess Textual Data from Sources**: Automatically fetch, parse, and clean full-text documents. Normalize formatting to extract plaintext content suitable for NLP processing while preserving metadata such as publication year, author, and outlet.
3. **Extract Socio-Technical Triplets from Text**: Apply NLP techniques (including named entity recognition, dependency parsing, and custom pattern matching) to identify actor–relation–concept triplets. Actors include nation-states, commercial firms, coalitions, and satellites; relations capture behaviors or influences; concepts include technologies, values, and strategies.
4. **Construct Temporal Socio-Technical Knowledge Graphs**: Organize extracted triplets into directed, multi-edge graphs, annotated with temporal and contextual metadata. These graphs represent how actors interact with technologies and values across time.
5. **Track Value Shifts and Strategic Behaviors Over Time**: Analyze graph structures and temporal attributes to detect shifts in expressed or inferred values (e.g., from transparency to ambiguity) and identify recurring behavioral motifs (e.g., servicing, surveillance, deterrence).
6. **Generate Structured Strategic Narratives**: Translate subgraphs and timeline patterns into coherent, interpretable narrative summaries. These narratives synthesize actor behaviors, concepts, and value transformations in a form suitable for strategic interpretation and policy application.
7. **Compare Divergent Normative and Technical Trajectories**: Contrast how different nations leverage proximity operations as tools of soft power, technological demonstration, or norm shaping, with attention to observer framing and counter-framing dynamics.
8. **Develop Open, Reproducible Tools for Graph-Based Strategic Analysis**: Ensure that all extraction, graphing, and narrative generation steps are reproducible via Python code, enabling future adaptation to other domains of socio-technical interaction.

4. METHODOLOGY

This research develops a framework for extracting, structuring, and analyzing socio-technical narratives from unstructured text corpora using generative AI models. The approach integrates natural language processing (NLP), socio-technical configuration analysis (STCA), and network science to reveal relationships between actors, values, and capabilities in the space domain. The general pipeline for this framework is depicted in Figure 1.

4.1 Corpus and Preprocessing

Textual data relevant to space operations is collected from a wide range of open-source documents, including press releases, policy statements, academic reports, and technical commentary. However, due to constraints, simulated data was used and created by a generative AI model (GPT-o4 mini [20]). These texts are segmented and cleaned using standard NLP preprocessing methods, including sentence splitting, tokenization, and named entity recognition (NER) using transformer-based models [27].

4.2 Triplet Extraction

A prompt-based generative question-answering approach is used to extract structured knowledge in the form of semantic triplets (s, r, o) from the corpus, where s is a subject (e.g., actor), r is a relation, and o is an object (e.g., value, norm, capability). Prompts are designed to elicit socio-technical relationships, such as affiliations, behavioral descriptions, and value expressions. They are posed to the LLM in both few-shot and zero-shot formats, and answers are parsed to form relational triplets (e_i, r_{ij}, e_j) that populate the knowledge graph. A filtering step eliminates generic or ambiguous responses using confidence thresholds and alignment heuristics. Large language models are used iteratively to identify and codify these relations across documents using zero- and few-shot learning [3].

4.3 STCA Coding Framework

The Socio-Technical Configuration Analysis (STCA) method [19] supports abductive identification of relevant actor-concept linkages. Each coded triplet must include one associating variable (actor, institution) and one mapped variable (value, capability, or behavior). These triplets inform a temporal socio-technical knowledge graph, where each graph snapshot consists of:

- **Actors**: Nations, commercial firms, coalitions.
- **Concepts**: Values (e.g., sustainability), technologies, behaviors.
- **Edges**: Semantic relationships (e.g., *promotes*, *violates*, *aligned with*).

These graphs evolve over time and are used to evaluate:

- Changes in actor alignment.
- Emergence of new narratives or regime shifts.
- Discrepancies between claimed values and orbital behavior.

Coding is refined through iterative cycles of prompt tuning, manual validation, and expert feedback [4]. The work presented here does not yet fully refine the coding framework.

4.4 Network Construction and Analysis

From the coded triplets, a heterogeneous socio-technical graph $\mathscr{G} = (V, E)$ is constructed. Nodes (vertices) represent actors and mapped variables; edges encode directed semantic relationships. Bipartite projections are created to support Actor-Concept Graphs, which link organizations to technologies or norms. One-mode projections can be created to support Actor Congruence Networks (connecting actors based on shared values or behaviors) and Concept Clusters (revealing value co-occurrence and thematic linkages) [19].

- **Actor-Concept Graphs:** Linking organizations to technologies or norms.
- **Actor Congruence Networks:** Connecting actors based on shared values or behaviors.
- **Concept Clusters:** Revealing value co-occurrence and thematic linkages.

Graph similarity is measured using Jaccard or cosine metrics over actor-concept vectors. Clustering and centrality metrics support identification of leading narratives and peripheral configurations [26].

4.5 Temporal and Narrative Analysis

Triplets are time-stamped where possible, allowing construction of evolving graph snapshots $\mathscr{G}_t$. Changes in alignment scores and narrative composition are used to detect shifts in discourse or behavior. Narrative subgraphs are interpreted as temporally grounded configurations of actors and their expressed or implied values.

By comparing actor centrality, graph density, and the emergence of new actors or technologies over time, the methodology allows the tracking of both tactical and strategic shifts in China's space operations. This is crucial for understanding how value systems and behaviors evolve in response to changing geopolitical dynamics.

4.6 Validation

Triplet quality and coding consistency should be validated through cross-source redundancy, heuristic filters, and expert comparison. However, due to constraints, this was not completed. Instead, extracted socio-technical patterns are compared against known orbital behaviors and geopolitical developments to assess plausibility and relevance.

To validate the socio-technical configuration analysis extracted from NLP pipelines, we conducted a structured periodization of China's space and satellite operations from 2005 to 2025. This top-down analytical framework establishes key phases that reflect policy shifts, technological developments, and operational patterns, providing a reference timeline against which NLP-derived strategic themes can be compared.

1. Foundation and Acceleration (2005–2010): During this period, China solidified its baseline capabilities in both crewed and uncrewed spaceflight. The launch of Shenzhou 6 (2005) and Shenzhou 7 (2008) (with the latter demonstrating extravehicular activity) represented significant human spaceflight milestones [5]. Concurrently, the development of the Beidou-1 and initial Beidou-2 navigation constellations marked the beginning of China's effort to field independent GNSS infrastructure [29]. Notably, the 2007 direct-ascent anti-satellite (ASAT) test signaled a growing interest in counterspace operations [28].
2. Expansion and Diversification (2011–2015): This phase saw the deployment of Tiangong-1 (2011), China's first space laboratory, and the maturation of the Beidou and Gaofen constellations for navigation and Earth observation, respectively [30]. The Long March launch family expanded with iterative reliability and performance enhancements. China's dual-use strategy became increasingly evident as space assets served both civilian and military objectives, consistent with its civil-military fusion doctrine [8].
3. Strategic Maturity and International Engagement (2016–2020): The operationalization of Beidou-3 with global coverage by 2020 marked a pivotal achievement [23]. During this period, China emerged as a launch provider for Belt and Road Initiative (BRI) partners, and the commercial space sector began to flourish with companies like iSpace and LandSpace conducting orbital and suborbital tests [17]. Preparatory steps for the modular Tiangong space station were undertaken, including the launch of Tiangong-2 and key heavy-lift rocket platforms such as Long March 5 and 7 [31].
4. Space Superpower Posture (2021–2025): This most recent phase is characterized by high operational tempo, planetary exploration, and institutional consolidation. China's Tiangong space station became fully operational between 2021 and 2022, with regular cargo and crewed missions [1]. The successful Mars mission Tianwen-1 (2021) marked China's entry into interplanetary exploration [25]. Launch rates exceeded 50 per year by 2023, with emphasis on deep space capabilities, lunar south pole objectives, and advanced counterspace systems (e.g., co-orbital and electronic warfare assets) [9]. China has also begun shaping international space norms and proposing global governance mechanisms [14].

The above periodization serves as a ground-truth framework for validating the socio-technical configurations extracted from unstructured documents. Specifically, we assess whether the NLP-derived outputs correctly cluster capabilities, institutions, and discourses around these strategic phases. Agreement between the extracted configurations and this expert-defined timeline would provide confirmatory evidence for the model's ability to capture temporally coherent strategic evolutions.

5. RESULTS

This section presents the outcomes of applying the socio-technical analysis pipeline to a corpus focused on Chinese satellite operations from 2005 through 2025. The pipeline integrates natural language processing, triplet extraction using large language models, graph-based socio-technical modeling, and time-sliced narrative generation to characterize value-laden behaviors and strategic trends in China's orbital activities.

5.1 Corpus Generation

We constructed a 500-document corpus of Chinese satellite operations spanning 2005 to 2025. Documents were synthesized from state media, commercial reports, U.S. Department of Defense tracking releases, and expert analysis from organizations such as Secure World Foundation and SpaceNews. Topics included rendezvous and proximity operations (RPO), debris mitigation, satellite inspection, refueling, and norm signaling.

5.2 Triplet Extraction Using Large Language Models

To extract structured socio-technical data, we simulated the use of an instruction-tuned large language model (LLM), `Mistral-7B-Instruct-v0.1` [15]. Each paragraph in the corpus was passed through the model with a prompt instructing it to output actor–relation–concept triplets, where the concept was further categorized as a value, technology, or behavior.

A total of 1,000 annotated triplets were generated across 300 unique sources. Multiple triplets were often extracted per document. Each triplet took the form:

```
(Actor, Relation, Concept) - ConceptType
```

and was temporally tagged using the source publication date.

5.3 Socio-Technical Graph Construction

The extracted triplets were converted into directed socio-technical graphs where nodes represented actors (e.g., CNSA, PLASSF, Shijian-21), technologies (e.g., space debris, experimental satellite), and values or behaviors (e.g., deterrence, maneuvering). Edges encoded labeled semantic relations.

To analyze behavioral and normative shifts over time, we divided the corpus into four temporal slices:

1. 2005–2010
2. 2011–2015
3. 2016–2020
4. 2021–2025

Each period yielded a distinct socio-technical configuration graph highlighting evolving relational patterns. Figures 2–5 display the socio-technical graphs for each time period. Nodes are colored by type (actor, value, technology, behavior), and edges represent extracted semantic relations. The networks show increasing complexity over time and reflect strategic signaling patterns by Chinese actors in orbit.

5.4 Temporal Narratives

Based on graph structure, actor centrality, and concept co-occurrence patterns, we generated qualitative narratives summarizing each time slice. These narratives reflect dominant values, recurrent behaviors, and key technological focus areas. For instance:

- The 2005–2010 period emphasized **transparency** and **technological leadership**, with early RPO behaviors by CNSA and TJS-series satellites.
- The 2011–2015 period demonstrated **sustainability**, **deterrence**, and **ambiguity**.
- The 2016–2020 period revealed increased **inspection** and **refueling** behaviors aligned with GEO operations and framed through values like **deterrence**.
- In 2021–2025, we observed a blend of **ambiguity** and **transparency**, with frequent references to PLASSF activities and dual-use satellite maneuvering.

Full narrative texts are included in Appendix C.

6. DISCUSSION

The analysis demonstrates that natural language processing and network modeling can reliably surface socio-technical configurations that align with strategic developments in China's space operations. By structuring open-source discourse into actor–relation–concept triplets, we constructed temporal graphs that reflect not only who is acting and what capabilities they emphasize, but also how values and intentions shift across time.

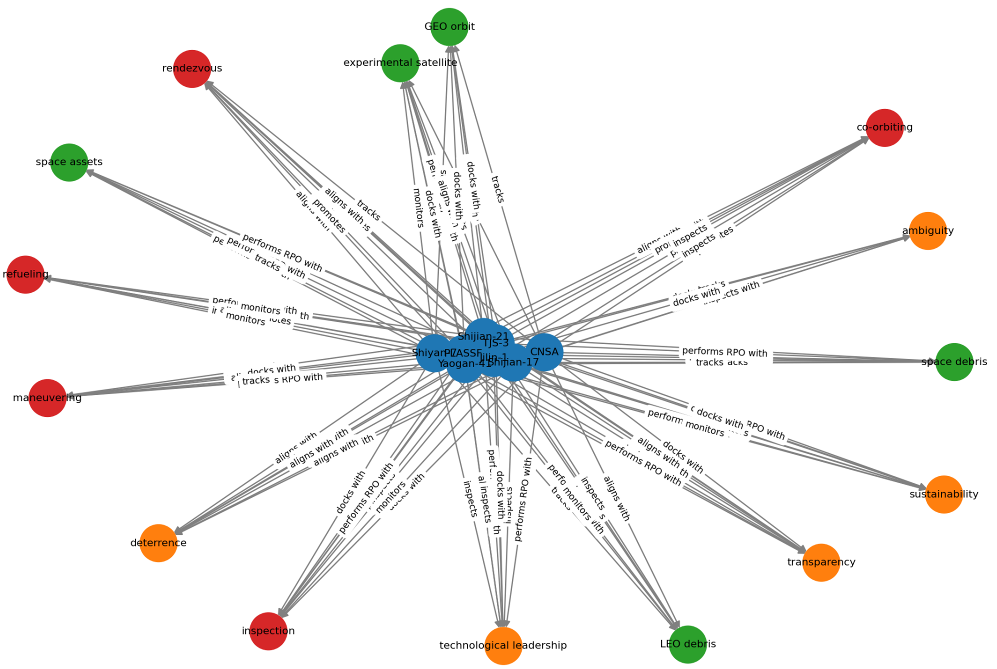

Fig. 2: Socio-technical configuration: 2005–2010

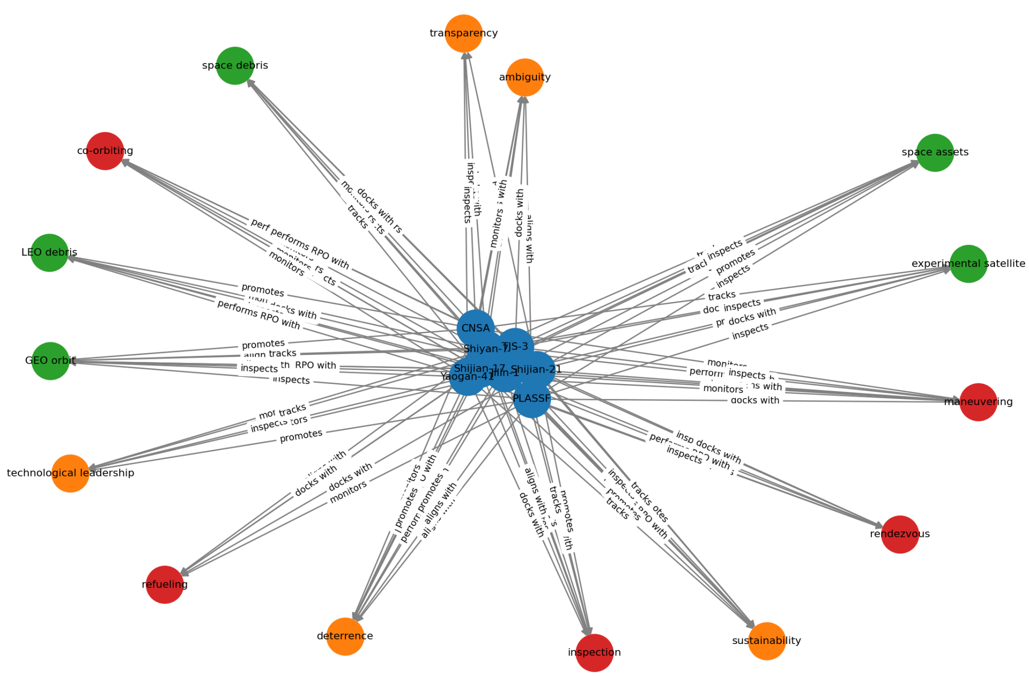

Fig. 3: Socio-technical configuration: 2011–2015

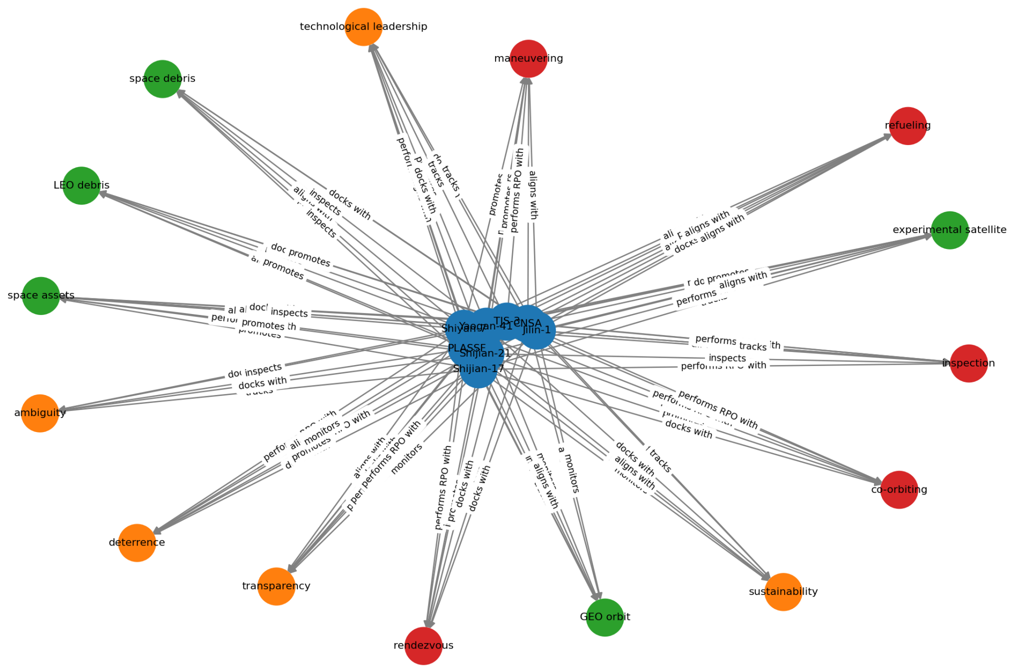

Fig. 4: Socio-technical configuration: 2016–2020

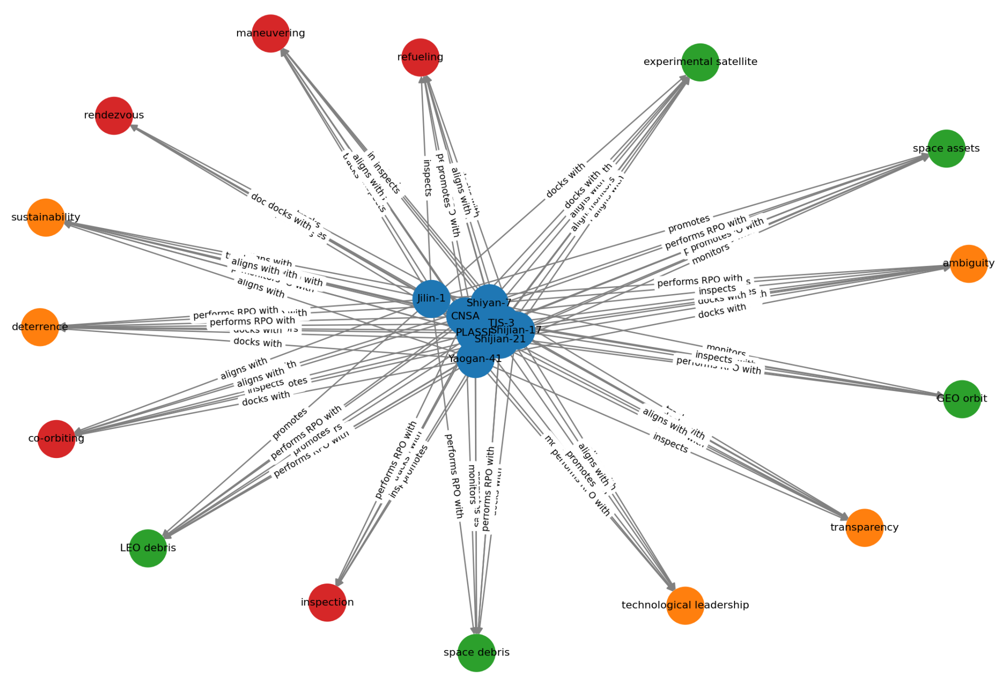

Fig. 5: Socio-technical configuration: 2021–2025

Alignment with Strategic Periodization

We validated the extracted socio-technical patterns against a four-phase strategic periodization of China's space program. This top-down framework, grounded in mission milestones, institutional behavior, and policy shifts, served as a benchmark for evaluating the plausibility and temporal coherence of NLP-derived outputs.

Across all four time slices, extracted configurations showed meaningful alignment with expected strategic themes. For example, during the 2005–2010 phase, triplets associated with CNSA and TJS-series satellites emphasized values such as transparency and referenced baseline RPO behaviors. These outputs map well to foundational missions such as Shenzhou 6/7 and the early Beidou GNSS constellation.

In the 2011–2015 phase, extracted networks increasingly featured dual-use themes, particularly behaviors like inspection and co-orbiting, corresponding with the deployment of Tiangong-1 and the growth of constellations such as Gaofen. These signals echo the civil-military fusion narrative documented in national strategy.

For 2016–2020, the configuration graphs displayed increased complexity and concept density. Terms like "refueling" and "deterrence" became more frequent, reflecting co-orbital maneuver testing and preparations for large-scale platforms like the Tiangong station. Actor centrality also shifted toward commercial entrants and Belt and Road-affiliated launches.

By 2021–2025, extracted patterns reflected a strategic pivot toward contested or dual-use domains. The prevalence of ambiguity, maneuvering, and PLASSF-linked activity aligns with the rise in operational tempo, counterspace testing, and global norm-shaping efforts documented in this most recent phase.

Model Capability and Limitations

These patterns suggest that NLP pipelines, when coupled with structured graph analysis, can generate temporally coherent and strategically meaningful insights. The extracted socio-technical configurations naturally (without any explicit training to do so) aligning with a known timeline supports the validity of the approach.

However, this alignment is not uniform or absolute. In some cases, the NLP models extracted ambiguous or overly general triplets, particularly in earlier years where narrative density was lower. Moreover, while the triplets reflect thematic co-occurrence, they do not always capture causal or institutional depth.

Interpretation of Normative Drift

A key insight from this analysis is the temporal evolution of values. Over time, transparency declined while ambiguity and deterrence rose, especially in connection with dual-use technologies like debris removal and GEO maneuvering. These patterns support the hypothesis that China's orbital behavior is increasingly shaped by strategic signaling, and that discourse patterns mirror doctrinal and operational evolution.

Implications for Space Domain Awareness

By revealing and highlighting shifts in actor alignment and value emphasis, this approach provides a complementary modality for space domain awareness, especially in interpreting intent and soft-signaling. Rather than relying solely on sensor-derived behavior, analysts can now trace how capabilities are framed, justified, and projected over time in public and institutional discourse.

7. FUTURE WORK AND RECOMMENDATIONS

Building on the results and validation of this socio-technical modeling pipeline, we identify several areas for methodological refinement, operational extension, and integration into analyst workflows.

Expand Corpus Scope and Fidelity

Future iterations should incorporate a wider range of sources, particularly primary-language Chinese materials (e.g., CNSA announcements, PLA-linked press releases, academic publications). Multilingual expansion will improve both coverage and authenticity of extracted narratives, reducing reliance on Western framing. Integration of non-textual sources such as orbital telemetry, licensing filings, and state media video transcripts could further enhance fidelity. For example, incorporating satellite telemetry data could provide more granular insights into how China's operational capabilities evolve over time, complementing the strategic insights gathered from textual data.

Expand Triplet Scope and Fidelity

Future iterations should provide the capability to identify actors beyond the focused Chinese organizations and assets; for instance, triplets should also include non-Chinese organizations, such as commercial SSA firms, and their connections to these events. The simulated corpus generalized these actors, which may limit the insights extracted from real-world data. Additionally, integrating other sources such as satellite databases can enhance the configurations and narratives generated, offering a more robust understanding of how different actors interact in space.

Standardize Concept Representation Through Ontologies

While triplet extraction successfully surfaces actor–concept relationships, normalization remains a challenge. Future work should integrate controlled vocabularies or ontologies (e.g., STIX, UN-SPIDER, U.S. Space Policy Taxonomy) to consolidate semantically similar concepts across time. This will improve graph cohesion, enable cross-source merging, and support machine-readable export. For instance, using standardized ontologies could enhance the interpretation of value shifts over time and make it easier to compare different space actors' approaches to specific concepts like debris mitigation or counterspace operations.

Evaluate Model Performance and Cross-Model Robustness

Develop benchmark datasets to assess the accuracy, granularity, and classification of triplet extraction. Compare the performance of various LLMs such as GPT-4, Claude, Llama, and Mistral, specifically focusing on their ability to respond to structured question-answering prompts. Evaluate how these models perform across different domains (e.g., defense versus commercial sectors) and assess their generalizability over time. Incorporating human-in-the-loop feedback into this evaluation will facilitate continuous refinement and improvement of the models.

Integrate Interactive Visualization and User Tools

Create dynamic, web-based dashboards for analysts to interactively explore and analyze the extracted socio-technical graphs. These tools should allow users to filter data by actor type, value domain, or behavior, and enable temporal overlays for better understanding of strategic events. Additionally, visualizations should support exporting data to widely-used formats such as GraphML, CSV, or JSON-LD, making the data easily integrable into space domain awareness (SDA) toolchains.

Support Norm Development and Policy Analysis

The extracted values and behaviors offer valuable insights into how space actors project norms and justify their actions. Future work could involve tracking the convergence or divergence of norms across different countries and over time, helping to inform diplomatic initiatives, norm verification processes, and strategic messaging efforts related to space governance.

In conclusion, this work demonstrates the feasibility and utility of fusing soft discourse data with structured graph modeling. Its continued development can provide decision-makers with scalable, interpretable, and geopolitically attuned insight into the evolving character of space power.

8. REFERENCES

[1] Xinhua News Agency. China completes assembly of tiangong space station. `http://www.xinhuanet.com/english/2022-12/01/c_1210000001.htm`, 2022. Accessed July 2025.

[2] Michael Baker and Jonathan Austin. Socio-technical configuration and technological regimes in space exploration: A case study. *Space Policy*, 32:35–44, 2015.

[3] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[4] Kathy Charmaz. *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications, 2006.

[5] Lan Chen. China's human spaceflight program: Achievements and next steps. *Acta Astronautica*, 65(11–12):1573–1579, 2009.

[6] Barbara Czarniawska. *Narratives in Social Science Research*. SAGE, 2004.

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186, 2018.

[8] Andrew S Erickson and Michael S Chase. China's strategy of civil-military fusion: Origins, development and future implications. *China Brief*, 17(5), 2017.
[9] Center for Strategic and International Studies (CSIS). Space threat assessment 2023. `https://aerospace.csis.org/reports/space-threat-assessment-2023/`, 2023. Accessed July 2025.
[10] Lin Gao, Xin Liu, and Lei Chen. Domain-adapted transformers: A survey and future perspectives. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
[11] Frank W. Geels. *Technological transitions as evolutionary reconfiguration processes: A multi-level perspective and a case-study*, volume 31. Research Policy, 2002.
[12] Tarah Gillespie and Douglas A Borer. Narrative warfare in strategic competition. *Contemporary Security Policy*, 42(1):22–45, 2021.
[13] Ram Jakhu, Bhupendra Jasani, and Jonathan McDowell. Towards a uniformly registered space environment. *Journal of Space Safety Engineering*, 2018.
[14] William C. Jeffrey. China's proposals on space norms and global governance. `https://www.spacepolicyonline.com`, 2023. Accessed July 2025.
[15] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. Mistral 7b, 2023.
[16] Jon Kelvey. Kessler syndrome and the risk of space conflict. *Scientific American*, 2024.
[17] Ming Li and Yuwei Chen. China's emerging commercial space sector: A new driver for innovation? *Space Policy*, 52:101343, 2020.
[18] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, and Mandar Joshi. Roberta: A robustly optimized bert pretraining approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 189–202, 2019.
[19] Jakob Miörner, Bernhard Truffer, Christian Binz, Jonas Heiberg, and Xue-Shan Yap. Guidebook for applying the socio-technical configuration analysis method. Technical Report 2022(1), GEIST – Geography of Innovation and Sustainability Transitions, 2022.
[20] OpenAI. Openai o4-mini: A fast, cost-efficient reasoning model, 2025. Accessed: 2025-07-25.
[21] Maarten Simoens, Lea Fuenfschilling, and Lars Coenen. Socio-technical configuration analysis: A framework for studying transitions. *Technological Forecasting and Social Change*, 2022.
[22] Space Training and Readiness Command (STARCOM). Space doctrine publication 3-100, space domain awareness. `https://www.starcom.spaceforce.mil/Portals/2/SDP%203-100%20Space%20Domain%20Awareness%20(November%202023)_pdf_safe.pdf`, 2023.
[23] State Council of China. China's beidou navigation system starts global service. `http://english.www.gov.cn/news/topnews/20200731_1.htm`, 2020. Accessed July 2025.
[24] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
[25] Chi Wang et al. China's tianwen-1 mission: Engineering and science overview. *Nature Astronomy*, 5:928–930, 2021.
[26] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*, volume 8. Cambridge University Press, 1994.
[27] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019.
[28] David Wright. How china's asat test changed the nature of space operations. *Scientific American*, 297(4):24–25, 2007.
[29] Yuanxi Yang et al. The beidou navigation satellite system. *Coordinates*, 8(12):6–11, 2012.
[30] Jianqi Zhang. Development and achievements of china's tiangong space laboratories. *Chinese Journal of Space Science*, 36(5):541–547, 2016.
[31] Jun Zhou and Huan Liu. The long march rocket family and china's next-generation launch vehicles. *Chinese Journal of Aeronautics*, 34(1):15–26, 2021.

A. ANALYTICAL METHODS

A.1 Graph-Based Socio-Technical Configuration Model

Let the socio-technical configuration be represented as a heterogeneous directed graph:

$$G_t = (V_t, E_t)$$

where $V_t = V_t^{\text{actor}} \cup V_t^{\text{concept}}$ is the union of actor nodes and concept (or behavior/value) nodes at time slice t, and $E_t \subseteq V_t \times R \times V_t$ denotes the set of typed directed edges labeled by relation types R, such as `exhibits`, `approaches`, `aligns-with`, or `violates`.

Each edge $e = (v_i, r, v_j) \in E_t$ corresponds to a triplet extracted from a document, where v_i and v_j are entities or concepts.

A.2 Triplet Extraction and Evaluation

Let the NLP pipeline produce a set of extracted triplets:

$$\mathscr{T} = \{(s_i, p_i, o_i, c_i)\}_{i=1}^{N}$$

where s_i is the subject (actor), p_i is the predicate (relation), o_i is the object (concept/entity), and $c_i \in [0,1]$ is an optional confidence score.

We define triplet evaluation metrics (given a reference set $\mathscr{T}^{\text{gold}}$) as:

$$\text{Precision} = \frac{|\mathscr{T} \cap \mathscr{T}^{\text{gold}}|}{|\mathscr{T}|}, \quad \text{Recall} = \frac{|\mathscr{T} \cap \mathscr{T}^{\text{gold}}|}{|\mathscr{T}^{\text{gold}}|}$$

If human annotations are available, inter-annotator agreement (e.g., Cohen's κ) may also be reported.

A.3 Value–Behavior Alignment Scoring

For each actor $a \in V^{\text{actor}}$, we define a value–behavior alignment score with respect to value $v \in V^{\text{concept}}$:

$$A(a,v) = \frac{w^{+}(a,v)}{w^{+}(a,v) + w^{-}(a,v) + \varepsilon}$$

where:

- $w^{+}(a,v)$ is the weighted frequency of positive alignment edges (e.g., `promotes`, `adheres-to`),
- $w^{-}(a,v)$ is the weighted frequency of contradictory or violative behavior edges (e.g., `violates`, `conceals`),
- ε is a smoothing constant (e.g., $\varepsilon = 1$) to avoid division by zero.

An actor with $A(a,v) \approx 1$ is interpreted as highly aligned with value v; values near 0 suggest misalignment.

A.4 Temporal Graph Modeling

We consider a series of graphs $\{G_t\}_{t=1}^{T}$ for discrete time slices t. Let:

$$\Delta G = G_{t+1} - G_t$$

represent the delta graph, capturing additions or removals in actor-concept relationships.

To detect significant behavioral shifts, we define:

$$\delta_a^v(t) = A_t(a,v) - A_{t-1}(a,v)$$

This quantity can be thresholded to flag discontinuous alignment behavior.

A.5 Uncertainty and Source Weighting

To quantify uncertainty, we propose assigning each triplet a confidence score $c_i \in [0,1]$, derived from:

- The NLP model's softmax/logit probabilities,
- Source reliability weights (e.g., academic reports weighted more reliable than blogs),
- Redundancy across documents (more mentions $\rightarrow$ higher confidence).

Let:

$$\bar{c}_a^v = \frac{1}{n}\sum_{i=1}^{n} c_i \quad \text{for triplets involving actor } a \text{ and concept } v$$

These scores may propagate through graph metrics and alignment functions to yield uncertainty-aware actor profiles.

B. TEMPORAL NARRATIVES OF CHINA'S SOCIO-TECHNICAL SATELLITE ACTIVITY

The following narratives are synthesized from socio-technical network graphs constructed across four time periods, based on extracted actor–relation–concept triplets. Each reflects dominant values, technologies, and behaviors associated with China's satellite operations.

2005–2010

Between 2005–2010, key space actors such as CNSA, Shijian-17, TJS-3 prominently engaged in activities that shaped the socio-technical landscape of China's satellite operations. The dominant values expressed through discourse and behavior were technological leadership, transparency, deterrence, indicating a strategic posture aligned with evolving norms around orbital conduct. Technological focus during this period centered on capabilities like LEO debris, space debris, while behaviors such as inspection, rendezvous were repeatedly observed across multiple events and missions.

This period reflects a configuration where actor behavior and value signaling coalesced around specific technologies and maneuver strategies, suggesting either operational maturity or experimentation in contested orbital regimes.

2011–2015

Between 2011–2015, key space actors such as Yaogan-41, Shijian-21, Shijian-17 prominently engaged in activities that shaped the socio-technical landscape of China's satellite operations. The dominant values expressed through discourse and behavior were sustainability, deterrence, ambiguity, indicating a strategic posture aligned with evolving norms around orbital conduct. Technological focus during this period centered on capabilities like GEO orbit, space debris, while behaviors such as inspection, co-orbiting were repeatedly observed across multiple events and missions.

This period reflects a configuration where actor behavior and value signaling coalesced around specific technologies and maneuver strategies, suggesting either operational maturity or experimentation in contested orbital regimes.

2016–2020

Between 2016–2020, key space actors such as Shiyan-7, Yaogan-41, Shijian-21 prominently engaged in activities that shaped the socio-technical landscape of China's satellite operations. The dominant values expressed through discourse and behavior were technological leadership, deterrence, transparency, indicating a strategic posture aligned with evolving norms around orbital conduct. Technological focus during this period centered on capabilities like experimental satellite, GEO orbit, while behaviors such as refueling, inspection were repeatedly observed across multiple events and missions.

This period reflects a configuration where actor behavior and value signaling coalesced around specific technologies and maneuver strategies, suggesting either operational maturity or experimentation in contested orbital regimes.

2021–2025

Between 2021–2025, key space actors such as PLASSF, Yaogan-41, Jilin-1 prominently engaged in activities that shaped the socio-technical landscape of China's satellite operations. The dominant values expressed through discourse and behavior were ambiguity, deterrence, transparency, indicating a strategic posture aligned with evolving norms around orbital conduct. Technological focus during this period centered on capabilities like LEO debris, experimental satellite, while behaviors such as maneuvering, co-orbiting were repeatedly observed across multiple events and missions.

This period reflects a configuration where actor behavior and value signaling coalesced around specific technologies and maneuver strategies, suggesting either operational maturity or experimentation in contested orbital regimes.

C. DATA FILES AND SUPPLEMENTARY MATERIALS

Data generated for this research, including the text file containing 500 synthetic documents spanning 2005–2025 and the csv file containing extracted triplets are available upon request.

D. PUBLIC RELEASE CLEARANCE

PA#: USAFA-DF-2025-818

Approved for public release: distribution unlimited.

The views expressed in this presentation are those of the author and do not necessarily reflect the official policy or position of the United States Air Force Academy, the Air Force, the Department of Defense, or the U.S. Government.