# Decoding Shifts in academic Networks: US-China Collaboration Dynamics in the GPT Era

Xi Lei*
xi.lei@epfl.ch
EPFL
Switzerland

Shu Yang†
shu.yang@epfl.ch
EPFL
Switzerland

Ruoxi Zhang‡
ruoxi.zhang@epfl.ch
EPFL
Switzerland

## Abstract

This study investigates how the emergence of large language models (LLMs), particularly GPT technologies, has reshaped the landscape of academic collaboration between the United States and China. Drawing on publication data from 2015 to 2025, we construct and analyze co-authorship, citation, and interdisciplinary networks among top-ranked institutions in both countries. We constructed institution-level and author-level collaboration networks using NetworkX and community-louvain for centrality and community detection analyses. We also built three-layer interdisciplinary co-occurrence networks based on OpenAlex's classification system. Our findings reveal a notable shift after 2020: international collaboration weakened while domestic ties within each country intensified. Interdisciplinary analysis further shows divergent trends in field convergence, with China emphasizing engineering and applied sciences, while the U.S. leans toward medicine and cognitive sciences. This research offers critical insights into how geopolitical shifts, strategic research agendas, and technological transformations impact global knowledge exchange for policymakers and academic stakeholders.

## 1 INTRODUCTION

Academic exchange and collaboration between China and the United States have long been a focal point for scholars. In recent years, with intensifying global technological competition and the rapid development of disruptive technologies, the significance and complexity of this collaborative relationship have become even more pronounced. Understanding and analyzing the evolution of academic cooperation between these two nations is crucial for grasping the trajectory of global technological advancement.

Existing research has extensively explored the dynamics of international scientific collaboration, primarily focusing on publication trends, citation impact, and co-authorship patterns across various fields and geopolitical contexts. For instance, studies have analyzed the publication output and influence of regions such as the European Union, United States, and China in top-cited papers, noting shifts in their relative contributions and the increasing embeddedness of China in international co-authored publications [6]. The impact of significant events, such as the COVID-19 pandemic, on the structure and internationalization of research teams, particularly between the U.S. and China in areas like coronavirus research, has also been a subject of investigation [2, 3]. Furthermore, analyses in specific high-impact fields like Artificial Intelligence have revealed patterns of talent migration and the benefits of cross-border collaboration between the U.S. and China, even amidst geopolitical tensions [1]. Broader examinations of Sino-American scientific copublications highlight the continuous rise of bilateral collaboration and its mutual benefits for both nations and global science [4, 5, 12].

Building on this foundational understanding, our research aims to investigate the specific influence of recent technological revolutions and global shifts on scientific collaboration between the U.S. and China. Previous scholarship has not yet fully examined the profound shifts in academic collaboration patterns specifically in the wake of transformative technological advancements, such as the rise of large language models. We seek to fill this gap by focusing on three key aspects of collaboration: co-authorship, citation networks (reflecting knowledge transfer), and interdisciplinary cooperation. To address the evolving landscape, we pose the following key research questions:

- How has the emergence of GPT technologies reshaped co-authorship and citation networks between Chinese and American research institutions?
- What significant differences can be observed in their collaboration paradigms, disciplinary structures, and knowledge flow characteristics before and after the rise of GPT models?

This study employs publication data from the OpenAlex database (2015-2025) in Computer Science, focusing on top U.S. and Chinese institutions. We construct and analyze institution-level and author-level co-authorship and citation networks using NetworkX and community-louvain for centrality metrics and community detection. Interdisciplinary co-occurrence networks are also built to explore cross-field collaboration. Our findings reveal a notable shift post-2020: international collaboration weakened, while domestic ties within both countries intensified. Chinese institutions gained centrality in collaboration and citation networks, indicating increased global influence. Interdisciplinary analysis shows divergent trends, with China emphasizing engineering and applied sciences, and the U.S. leaning towards medicine and cognitive sciences. This research underscores the impact of geopolitical shifts, strategic research agendas, and technological transformations on global knowledge exchange.

## 2 DATASET

We use the OpenAlex dataset to analyze academic publications in computer science from 2015 to 2025[8], focusing on leading institutions in China and the United States. Our selection includes papers from top U.S. universities (CS Ranking Top 30) and Chinese universities involved in the 985 project (Top 30). We exclude retracted publications and non-research content such as editorials or

*Both authors contributed equally to this research.
†Both authors contributed equally to this research.
‡Both authors contributed equally to this research.

metadata-only records. Only impactful papers are included, defined as those with a citation percentile value greater than 0.5.

## 2.1 Data Preprocessing

We focused on computer science publications affiliated with top institutions in the United States and China. Specifically, we included papers from the top 30 U.S. universities based on computer science rankings and Chinese universities involved in the 985 Project (also limited to the top 30). The data spanned the years 2015 to 2025. We excluded retracted publications and non-research content such as editorials or metadata-only records. To ensure quality, we retained only impactful papers with a citation percentile value greater than 0.5, which indicated that a paper was cited more often than at least 50% of other papers in the same field and year. After filtering, the final dataset included 30,347 papers from the U.S. and 52,994 papers from China, with 60.59% and 57.19% of the papers respectively classified as impactful. We also selected the Top 10 and Top 300 scholars based on total citation counts during the study period and used them as the focal groups for subsequent analysis.

# 3 METHODS

## 3.1 Network Construction

*Collaboration Network.* We first construct institution-level collaboration networks. For each paper, we match each author to their affiliated institution, retaining only authors affiliated with the top 30 and top 10 institutions (as defined in Chapter 2). Using NetworkX, we build graphs where nodes represent institutions and edges represent the number of co-authored publications between them. In total, we construct four institution-level networks: one each for the top 30 and top 10 institutions, both before and after 2020, as shown in Table 2.

Next, we shift our focus to author-level networks, specifically for the top 300 scholars (as defined in Chapter 2). We use NetworkX to build two graphs—one for the period before 2020 and one for after 2020—where nodes represent individual scholars and edges represent the number of co-authored papers, as shown in Table 2. To analyze patterns of scholarly clustering, we apply the Girvan–Newman algorithm[7], which detects community structure by progressively removing edges with the highest betweenness centrality.

**Table 1: Network Summary**

| Network | #Nodes | #Edges |
|---|---|---|
| `Top30 Institution Before 2020` | 58 | 1044 |
| `Top30 Institution After 2020` | 58 | 1090 |
| `Top10 Institution Before 2020` | 18 | 140 |
| `Top10 Institution After 2020` | 18 | 143 |
| `Top300 Scholars Before 2020` | 299 | 5949 |
| `Top300 Scholars After 2020` | 291 | 3193 |

*Citation Network.* We construct citation networks at the institutional level, resulting in four distinct networks: Top 30 and Top 10 institutions, each for the periods before and after 2020. Using NetworkX, we build graphs where nodes represent institutions and edges represent citation counts between them, as shown in Table 2. There are two types of undirected edges: citations from Chinese institutions to other Chinese institutions, from U.S. institutions to other U.S. institutions, and two types of directed edges: from Chinese to U.S. institutions, and from U.S. to Chinese institutions.

**Table 2: Network Summary**

| Network | #Nodes | #Edges |
|---|---|---|
| `Top30 Institution Before 2020` | 60 | 2808 |
| `Top30 Institution After 2020` | 60 | 3214 |
| `Top10 Institution Before 2020` | 20 | 414 |
| `Top10 Institution After 2020` | 20 | 418 |

*Interdisciplinary Network.* We construct co-occurrence networks to analyze Computer Science collaborations with other fields using OpenAlex's automated field classification system. Each paper can be tagged with multiple hierarchical fields (Level 0: major fields; Level 1: subfields). We used OpenAI's API with manual verification to map Level 1 subfields to Level 0 categories.

For papers from top 10 US and Chinese institutions, we built three-layer networks: Layer 1 (Computer Science), Layer 2 (co-occurring disciplines), Layer 3 (institutions). Edges connect CS to disciplines (weighted by co-occurrence papers) and disciplines to institutions (weighted by institutional papers in each discipline).

We analyze collaboration patterns across two time periods and regions, calculating co-occurrence ratios as CS-X collaborative papers divided by total CS papers. Sankey diagrams visualize the networks, with ratio differences between periods indicating collaboration trends (positive = increasing, negative = decreasing).

**Table 3: Interdisciplinary Network Paper Statistics**

| Region | Time Period | CS-X Papers |
|---|---|---|
| China | Before 2020 | 169,000 |
| China | After 2020 | 257,000 |
| US | Before 2020 | 115,000 |
| US | After 2020 | 98,000 |

## 3.2 Network Metrics

To understand how collaboration and citation patterns between research affiliations have changed over time, we applied a set of network analysis methods using Python's NetworkX and community-louvain libraries. We looked at two types of networks—co-authorship and citation—and for each, we analyzed two time periods: before 2020 and after 2020.

To find out which affiliations played the most central roles, we calculated three commonly used centrality metrics[13]. Degree centrality tells us how many direct links a node has, which shows how active or connected an affiliation is. Closeness centrality measures how close a node is to all others in the network, helping us see which affiliations could quickly reach or influence others. Betweenness centrality identifies nodes that often lie on the shortest paths between others, suggesting a kind of bridging or gatekeeping role. To get a more comprehensive view of each affiliation's overall importance, we also created an aggregate centrality score by normalizing and averaging these three metrics. This combined measure helps us identify institutions that excel across multiple dimensions of network influence, rather than just in one particular aspect.

To detect communities—groups of affiliations that are more connected to each other than to the rest of the network—we used the Louvain algorithm[9]. It groups nodes by maximizing modularity, and we visualized the results using the top nodes (ranked by degree centrality). Each community was shown in a different color, using a force-directed layout to highlight the relationships between them.

We also looked at how locally clustered the networks were by computing the clustering coefficient for each node[11]. This measures how likely it is that a node's neighbors are also connected to each other, which helps identify tightly connected groups. We plotted the overall distribution of these values for both time periods and then zoomed in on the 10 most clustered nodes to visualize their local networks.

## 3.3 Sankey Diagrams Construction

To analyze interdisciplinary collaboration patterns, we construct co-occurrence networks based on OpenAlex's hierarchical field classification system. Specifically, we focus on the top 10 U.S. and Chinese institutions and build three-layer networks representing the interaction between institutions and different disciplinary fields. This involves mapping subfields to broader disciplines, leveraging OpenAI's API for consistent categorization where necessary. We then construct Sankey diagrams to visually represent the flow of collaboration and knowledge between these layers[10]. These networks allow us to visualize and quantify the extent to which computer science research from these institutions integrates with and contributes to various other scientific domains, revealing shifts in interdisciplinary focus over time.

# 4 RESULTS

## 4.1 Collaboration Network

*1. Geographic Distribution.* This section analyzes the geographic evolution of the US–China academic collaboration network. We highlight shifts in regional concentration and cross-national link dominance.

**Cross–national dominance:** US–China collaborative links outnumber cross–country links among elite institutions.

**Major collaboration hubs:** Beijing, Shanghai, the Boston–Washington Corridor, and the San Francisco Bay Area emerge as key academic collaboration centers. Scholars from these regions engage in extensive partnerships both domestically and internationally, forming the backbone of the US–China research network.

**Major US–China Collaboration Links** Cross–national collaboration between the US and China was concentrated in a few key routes, particularly: Boston – Shanghai, Los Angeles – Beijing, Boston – Beijing.

**Rising North–South Collaborations Within China:** After 2020, there is a noticeable increase in north–south collaborations within China, indicating a more decentralized domestic collaboration pattern.

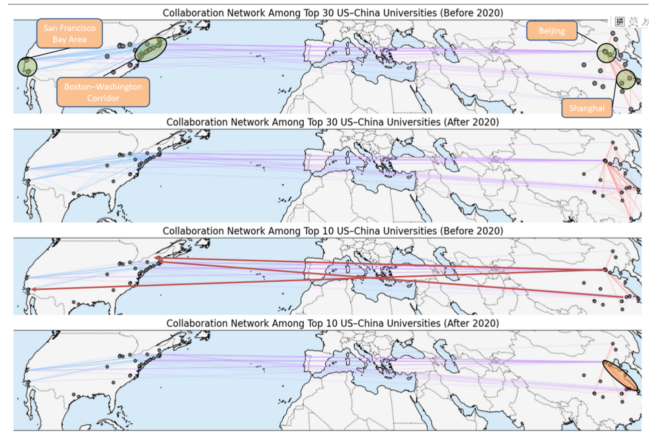


**Figure 1: Collaboration Network Among Top 30 and Top 10 US–China Universities Before and After 2020**

*2. Collaboration Network Shifts.* This section analyzes structural changes in the US–China collaboration network after 2020, focusing on cross–national decline and diverging domestic patterns. To highlight meaningful relationships, we filter out weak ties by excluding edges representing fewer than 10 co-authored publications.

**Decline in international collaboration:** After 2020, cross–national ties weaken significantly, and top universities such as Tsinghua, SJTU, Stanford, and UC Berkeley show noticeably thinner external links, which means deep collaborations among top–tier universities are diminishing.

**Diverging domestic trends:** While internal collaboration among Chinese universities intensifies after 2020, the internal ties within US universities show a slight decline.

**Limited ties for defense–linked Chinese universities:** Institutions such as HIT, UESTC, and NUDT maintain minimal international collaborations, consistent with national security considerations.

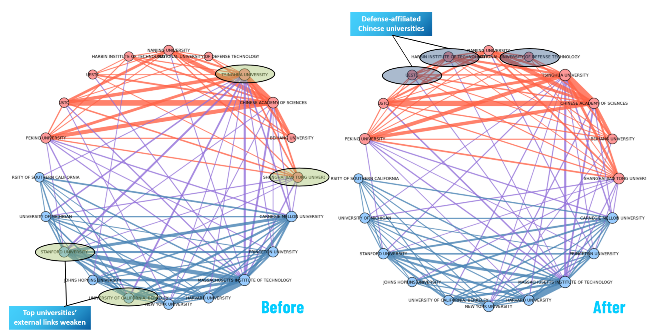


**Figure 2: Collaboration Network Among Top 10 US–China Universities Before and After 2020**

*3. Co–authorship Clustering.* This network represents co–authorship relationships among the top 300 scholars. Nodes correspond to individual scholars, with node size indicating the number of collaborators, and edge thickness reflecting the number of co–authored papers. Node colors denote citation–based ranking: red for the Top 50 scholars, orange for ranks 51–100, and blue for ranks 101–300.

Community structure is revealed via the Girvan–Newman algorithm.

**Before 2020: Core-Dominated Structure**

**Top Community: Densely Connected Core** The top community formed a tightly knit core network with near–complete internal connectivity.Top 50 scholars typically had fewer collaborators but produced high–impact publications, scholars ranked 101–300 adopted a broad–collaboration strategy, co–authoring extensively to build influence.

**Bottom Community: Loosely Structured & Dispersed** The bottom community is more fragmented and decentralized, exhibiting weaker internal cohesion. Top 50 scholars (red nodes) are dispersed across the periphery rather than forming a single tight block. This structure is characterized by bridge–builders and cross–institutional cooperation, likely reflecting a more international and mobile collaboration style.

**After 2020: Decentralized Expansion and Dispersion**

After 2020, the overall co–authorship network becomes denser but also more decentralized, marked by the emergence of multiple local clusters rather than a dominant central core. Top 50 scholars are more dispersed and no unified leadership group among top–tier scholars. In contrast, scholars ranked 101–300 have become more active and central, often serving as key connectors across different sub–communities. Additionally, a growing number of isolated or weakly connected nodes appear at the edges, likely reflecting the entry of new or interdisciplinary researchers who are still integrating into the main academic network.

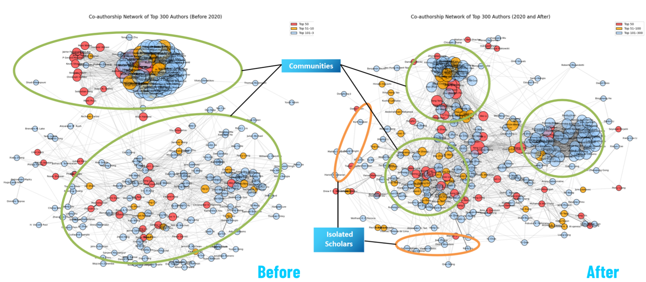


**Figure 3: Co-authorship Network of Top 300 Authors**

*4. Network Metrics.*

*4.1.1 Centrality Measures.* In our analysis, we examined all three centrality measures - degree, closeness, and betweenness - separately to understand different aspects of institutional influence. However, for clarity in presenting the main findings, we focus here on the aggregate scores that combine these dimensions. The full breakdown of individual metrics, along with their specific interpretations, can be found in the Appendix.

Global research collaboration has undergone a noticeable transformation when comparing the pre- and post-2020 periods (See Fig. 4, 5). The data tells an interesting story - Chinese universities have been climbing up the rankings in terms of their network influence, while some traditional U.S. powerhouses, though still strong, aren't quite as dominant as before.

What's particularly striking is how Tsinghua University has emerged as the new global leader in research connectivity after 2020. It's not just Tsinghua either - we're seeing more Chinese universities break into the top tier. On the other side of the Pacific, institutions like MIT and Harvard are still major players, but the numbers show they've lost some ground relatively speaking.

Also, the big picture shows research networks becoming less centralized around a couple of super-hubs. Where before you had a clear hierarchy with CAS and MIT at the very top, now we're seeing a more balanced distribution of influence, with multiple strong nodes from both China and the U.S. meaning global science is becoming less about one or two dominant centers and more about a network of equally important hubs working together.

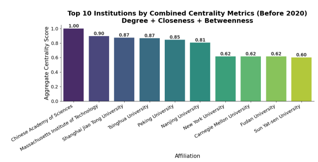


**Figure 4: Aggregate Cooperation Centrality of Top 10 Research Institutions (Before 2020)**

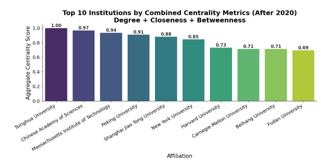


**Figure 5: Aggregate Cooperation Centrality of Top 10 Research Institutions (After 2020)**

*4.1.2 Louvain Community Detection.* The Louvain community detection results show a relatively clear split between two main communities in both periods (See Fig. 6, 7). Before 2020, the division appears more mixed, with Chinese and U.S. institutions distributed across both communities. After 2020, however, the structure becomes more polarized: Chinese institutions like Tsinghua University, Peking University, and Chinese Academy of Sciences are grouped more tightly together in one community, while most U.S. institutions form the other.

This may suggest that national-level collaboration has become stronger within each region. Chinese affiliations, in particular, show more internal connectivity after 2020, forming a more cohesive cluster. While international links are still present, the network seems to reflect more regionally concentrated collaboration patterns compared to the earlier period.

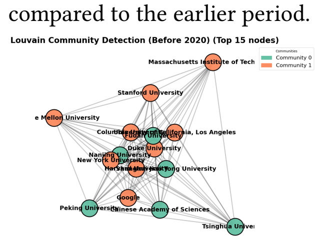


**Figure 6: Louvain Community Structure of Collaboration Network (Before 2020)**

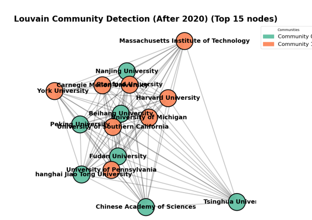


**Figure 7: Louvain Community Structure of Collaboration Network (After 2020)**

*4.1.3 Clustering Coefficient.* The clustering coefficient distribution shows that, after 2020, values became slightly more spread out, with more institutions showing very high local clustering. This means that, in the new period, some groups of universities became more tightly connected internally, forming small but strong clusters (see Fig. 17).

The subgraph of high-clustering nodes confirms this trend. In both periods, we can see small cliques of tightly linked institutions, but after 2020, the network appears slightly more dense and diverse in structure (see Fig. 9, 10). Chinese institutions also appear more often in the post-2020 subgraph, suggesting stronger regional collaboration within high-density groups.

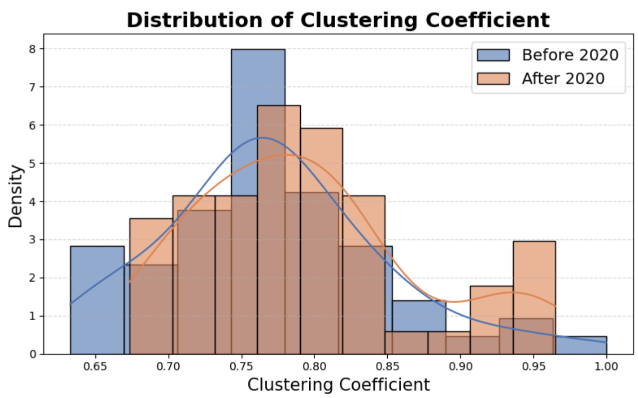


**Figure 8: Distribution of Clustering Coefficient Before and After 2020**

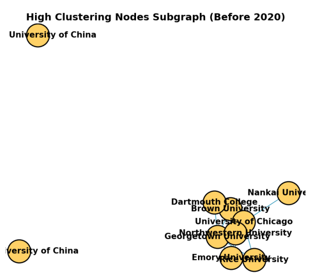


**Figure 9: Subgraph of High Clustering Nodes in Collaboration Network (Before 2020)**

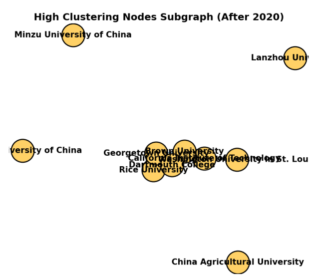


**Figure 10: Subgraph of High Clustering Nodes in Collaboration Network (After 2020)**

## 4.2 Citation Network

*1. Geographic Distribution.* This section analyzes the geographic evolution of the US–China academic citation network.

**Increase in Citation Links**

Edges in the citation network become noticeably thicker after 2020. This is expected, as publications from earlier years continue to accumulate citations over time, leading to stronger measured connections between institutions.

**Geographic Citation Clusters**

Distinct geographic citation clusters emerge on both sides of the US–China academic network. In the United States, high–density regions include the East Coast (e.g., Boston, New York), the Pittsburgh area, the West Coast (e.g., San Francisco Bay Area, Los Angeles), and a growing Southern hub (e.g., Texas). In China, cities such as Beijing, Shanghai, Guangzhou, and Harbin stand out as major citation centers, indicating concentrated academic activity and influence.

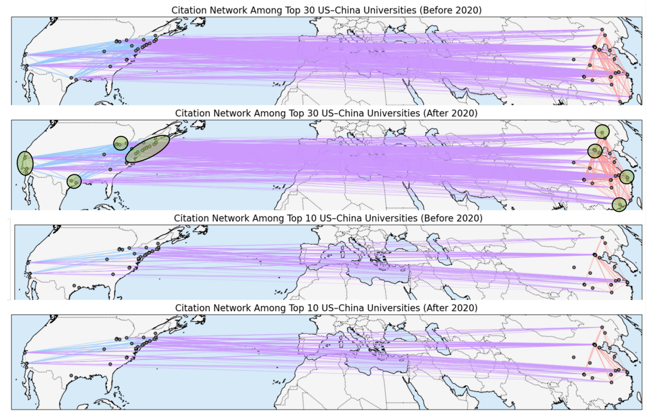


**Figure 11: Citation Network Among Top 30 and Top 10 US–China Universities Before and After 2020**

*2. Collaboration Network Shifts.* This section analyzes structural changes in the US–China citation network after 2020, focusing on citation direction, internal versus external citation patterns, and institutional influence. Since citations are unidirectional, we define four types of edges to distinguish citation flows: within Chinese universities (red edges), within US universities (blue edges), from China to the US (purple edges), and from the US to China (orange edges).

**Citation Flow Trends**

China's self-citation has increased, while internal citations among US universities have declined. This is reflected in denser red edges within Chinese institutions and lighter blue edges among US schools. Before 2020, citation flows primarily moved from China to the US. After 2020, however, US universities increasingly cite Chinese research–signaling growing academic recognition and rising research quality in China.

**Leading Institutions in Citation Influence**

Tsinghua University stands out as China's most cited institution, frequently referenced by both domestic and international peers. In the United States, MIT, Carnegie Mellon University (CMU), and Stanford University maintain high citation counts across national and global networks, consistently occupying central positions in the international citation landscape.

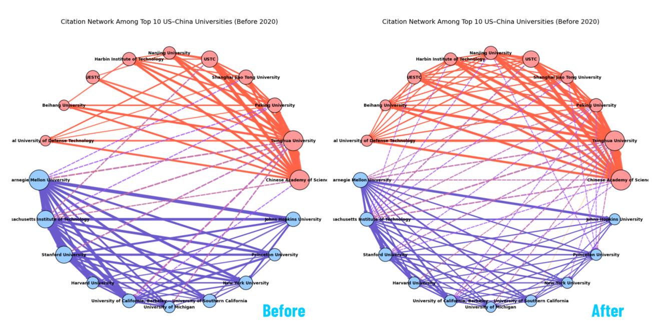


**Figure 12: Citation Network Among Top 10 US–China Universities Before and After 2020**

*4. Network Metrics.*

*4.2.1 Centrality Measures.* Apparently, the aggregated centrality scores show that institutions' prominence is showing a major shift post 2020. In the ranking post 2020, Chinese institutions dominate the top places, with a perfect score of aggregated centrality score achieved by the Chinese Academy of Sciences, Shanghai Jiao Tong University, Tsinghua University, and Fudan University. Conversely, before 2020, though Tsinghua University also dominated, on the whole there had been more even coverage with universities from the U.S. such as MIT, Carnegie Mellon University, and Stanford University at the top. This shows a solidifying centrality and Chinese institutions' dominance in the citation network, particularly in the recent years.

Since the network is constructed solely based on citation relationships, the observed increase in aggregate centrality scores after 2020 suggests that citation patterns among top institutions have become more concentrated and interconnected.

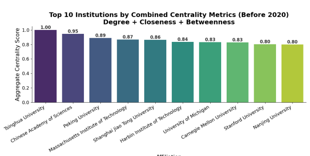


**Figure 13: Aggregate Citation Centrality of Top 10 Research Institutions (Before 2020)**

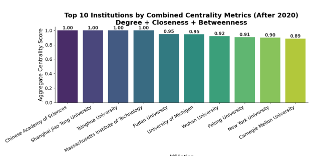


**Figure 14: Aggregate Citation Centrality of Top 10 Research Institutions (After 2020)**

*4.2.2 Louvain Community Detection.* Following the trends we saw in the centrality results, the Louvain community detection results give us another view of how the structure of the citation network has evolved. Before 2020, the top institutions were more scattered into several smaller and mixed communities (See Fig.15, 16). We can see Chinese and American affiliations spread across different clusters, without a very strong national boundary. However, after 2020, the community structure becomes more cohesive. Chinese institutions like Tsinghua, Peking, and Shanghai Jiao Tong form a clearer group, while U.S. universities like MIT, Berkeley, and Carnegie Mellon appear more closely grouped together. This shift suggests that national or regional clustering became more pronounced after 2020. Even though collaborations still exist across communities, the boundaries look more structured, which might reflect changes in international academic relationships.

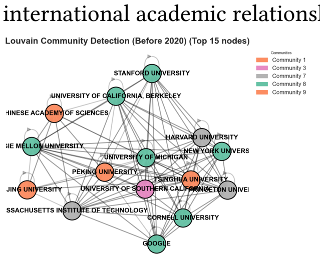


**Figure 15: Louvain Community Structure of Citation Network (Before 2020)**

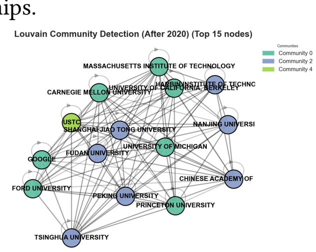


**Figure 16: Louvain Community Structure of Citation Network (After 2020)**

*4.2.3 Clustering Coefficient.* Finally, the clustering analysis helps us understand how tightly connected the institutions are. Looking at the distribution of clustering coefficients, we can see a clear shift after 2020—more institutions now have higher clustering values. This suggests that local collaboration patterns became stronger and more dense. The subgraph of high clustering nodes shows this change too. While before 2020 the highly clustered institutions were mostly American, after 2020 we see more Chinese universities showing up in this group. It means that China not only gained central positions in the network, but also developed more cohesive academic circles.

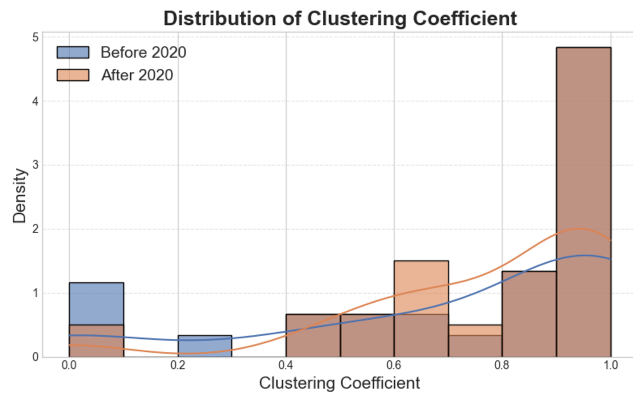


**Figure 17: Distribution of Clustering Coefficient Before and After 2020**

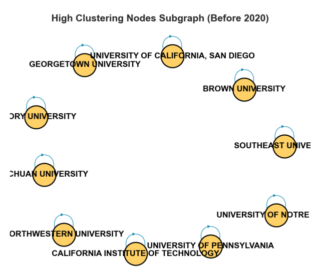


**Figure 18: Subgraph of High Clustering Nodes in Citation Network (Before 2020)**

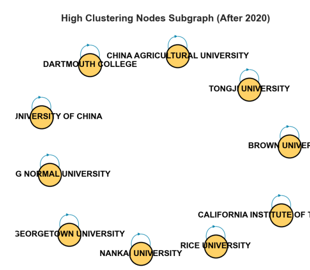


**Figure 19: Subgraph of High Clustering Nodes in Citation Network (After 2020)**

## 4.3 Interdisciplinary Network

*4.3.1 Field Collaboration Patterns.* This section looks at how Computer Science works with other academic fields. We found some surprising results that go against what we expected about STEM field relationships.

*Top Collaborating Fields Discovery.* We found different patterns in how Computer Science collaborates with other fields in China and the US. In China, the top 10 fields working with Computer Science are: (1) Computer Science, (2) Mathematics, (3) Physics, (4) Engineering, (5) Chemistry, (6) Psychology, (7) Philosophy, (8) Biology, (9) Political Science, and (10) Economics. The collaboration rates range from 40.1% to 43.3%, as shown in Fig. 20.

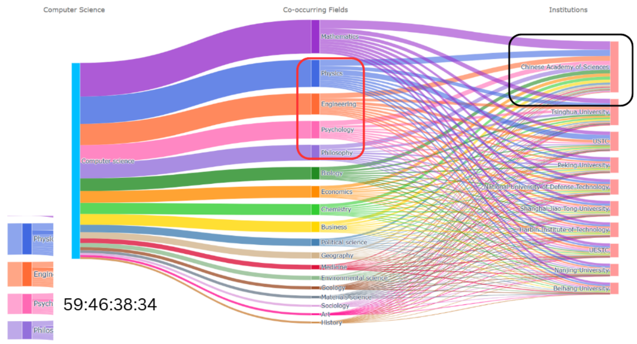


**Figure 20: CN Top10 Field**

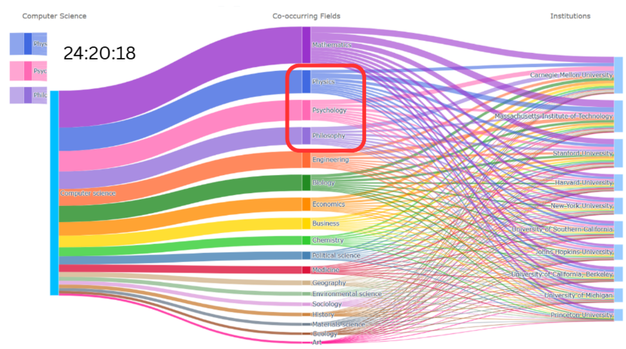


**Figure 21: US Top10 Field**

In contrast, in the US, the top 10 collaborating fields are: (1) Computer Science (41%), (2) Mathematics, (3) Physics, (4) Psychology, (5) Philosophy, (6) Biology, (7) Engineering, (8) Chemistry, (9) Economics, and (10) Medicine. The rates in the US range from 40.0% to 41.4%,as shown in Fig. 21.

*Deep Dive: Philosophy & Psychology Collaborations.* As shown in Fig. 20 and Fig. 21, Psychology and Philosophy are more important than we thought.

We originally expected traditional STEM fields (Mathematics, Physics, Engineering) to dominate Computer Science partnerships, but both Psychology and Philosophy rank in the top 6 for China and the US—highlighting the growing role of human–computer interaction, AI ethics, and cognitive computing.

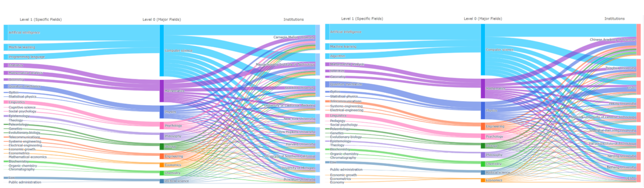

**Figure 22: CN US SubFieldDistribution.**

In Fig. 22, we take a closer look at each field's top three subdisciplines explains this shift:

- **Philosophy (Epistemology):** In the US, epistemology drives the majority of CS–Philosophy co-publications, reflecting strong engagement with knowledge representation and AI foundations. In China, Philosophy ties into CS more via application-oriented areas like HCI and cognitive ergonomics.
- **Psychology (Linguistics):** Linguistics dominates CS–Psychology work in the US, underscoring NLP and computational psycholinguistics. Chinese CS–Psychology collaborations skew toward engineering applications such as user-interface design and cognitive workload modeling.

*Differences Between China and the US:.* Both countries have similar top fields, but there are some interesting differences. China focuses more on traditional engineering and physical sciences - Chemistry is 5th in China but only 8th in the US. The US seems to adopt interdisciplinary fields earlier - Psychology is 4th in the US but 6th in China. Also, medicine is in the top 10, but China has Political Science instead. These differences show that the two countries have different research priorities and strengths.

*China's Concentrated Pattern.* China exhibits a highly centralized structure dominated by the Chinese Academy of Sciences, followed by Tsinghua University, University of Chinese Academy of Sciences, University of Science and Technology of China, and National University of Defense Technology (Fig. 20). These national-level institutions focus on fundamental science, engineering applications, and high-tech development, forming a backbone-driven knowledge production model.

*US's Distributed Pattern.* The US demonstrates a decentralized pattern with collaboration more evenly distributed across institutions (Fig. 21). No single institution dominates like CAS in China, reflecting the US's distributed academic system where multiple universities contribute to interdisciplinary research.

China's concentration enables easier coordination but may limit diversity, while the US's distribution fosters innovation through competition but complicates large-scale coordination. Notably, entropy metrics (Figure **??**) measure *within*-institution field diversity, whereas Sankey diagrams show *between*-institution collaboration volume distribution. Chinese institutions are individually more diversified (higher entropy) but aggregate CS partnerships are centralized, while US institutions focus more narrowly (lower entropy) but distribute collaboration load more evenly.

*Institutional Strengths Drive Collaboration Patterns.* One reason some institutions dominate particular flows in the Sankey diagram is that they are themselves national (or global) leaders in those fields—so CS naturally "follows the strength." Two illustrative examples:

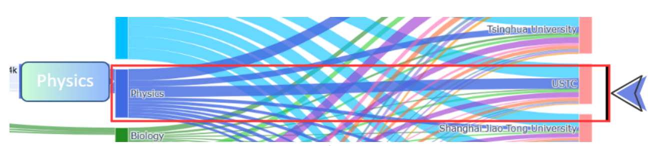


**Figure 23: Physics-USTC**

- **Physics → USTC:** The University of Science and Technology of China is widely recognized as China's top physics powerhouse. As a result, its share of CS–Physics co-publications (the thick purple ribbon in Fig. 23) is disproportionately large compared to other fields.
- **Medicine → Harvard:** Harvard University consistently ranks first globally in Medicine. This leadership draws Computer Science researchers into medical AI, computational

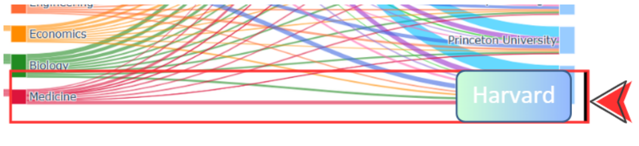


**Figure 24: Harvard-Medicine**

pathology, and bioinformatics partnerships—visible as the prominent red ribbon toward Harvard in Fig. 24.

These cases show that pre-existing institutional excellence in a subfield acts as a magnet for CS collaborations, reinforcing and amplifying established research strengths.

*4.3.2 Blue Ocean Opportunities.* This section identifies fields with minimal CS collaboration, representing potential growth areas.

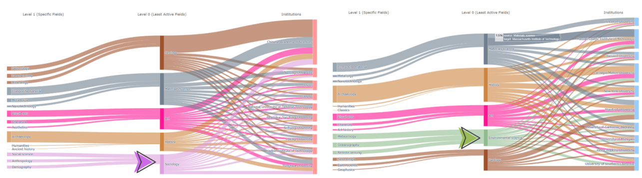

**Figure 25: The 5 Least Field in CN and US**

*Common Weak Fields.* Both countries show minimal CS collaboration with Geology, Materials Science, Art, and History (Fig. 25). These "blue ocean" opportunities could yield significant impact due to limited existing work.

*Country-Specific Gaps.* China lacks CS-Sociology collaboration, while the US shows weak CS-Environmental Science connections. These gaps reflect different national priorities and could offer innovative research opportunities—CS-Sociology in China for social media analysis, CS-Environmental Science in the US for climate modeling.

*4.3.3 Temporal Evolution.* We compared pre-2020 and post-2020 periods to identify changing collaboration patterns. The top five decreasing and increasing co-occurrence ratios are presented in Appendix 4.

Both countries show declining Mathematics collaboration (−0.089), along with Philosophy and Economics. Psychology collaboration increased in both regions, suggesting a shift toward human-centered research. China shows movement toward applied fields (Chemistry, Engineering, Materials Science), while the US emphasizes medical-engineering intersections (Medicine, Physics, Geography).

## 5 LIMITATION AND FUTURE WORK

While this study offers a multi-faceted view of changes in global academic collaboration, particularly between Chinese and US institutions, it faces several limitations. First, our analysis relies on OpenAlex's automated concept labels, which may introduce algorithmic bias affecting discipline categorization and influencing interpretations of collaboration patterns.

Second, although we observe notable structural shifts, such as increased domestic clustering and a decline in China-US co-authorship, we cannot confidently attribute these trends to specific causes like policy reforms or geopolitical tensions. Our findings are descriptive, requiring additional sources for causal understanding.

Another limitation is the dataset's scope, focusing solely on the top 30 institutions in China and the US. This exclusion of smaller or emerging research players and entire regions (Europe, South America, Africa) may mask important broader patterns of international cooperation. Future studies should aim for wider institutional and geographic inclusion.

Finally, while we initially assumed AI tools like GPT might promote greater cross-border collaboration, our observed patterns of strengthening national clusters do not yet support this. It remains unclear if these tools are reinforcing existing gaps or if their full impact is yet to be seen, requiring further investigation combining network data with qualitative insights.

## 6 CONCLUSION

To sum up, our analysis shows a notable shift in collaboration patterns between US and Chinese universities. Before 2020, cross-national ties—especially among top-tier institutions—were strong. However, after 2020, these links started to weaken. Domestic collaborations within each country intensified, while international ones, particularly between the US and China, declined. At the same time, the co-authorship network became denser and more decentralized. Top scholars became more evenly spread out, and we observed the emergence of more peripheral or weakly connected nodes. Several external factors likely contributed to this shift: increasing geopolitical tensions, a strategic move toward research self-reliance in both countries, and reduced academic mobility due to COVID-19. These conditions created structural barriers to sustained international collaboration.

We also observed a shift in the direction and structure of academic citation flows. Before 2020, citations largely moved from Chinese papers to US publications. After 2020, this pattern became more balanced and reciprocal, with growing international recognition for Chinese research. Nevertheless, the US continues to hold structural advantages, such as strong publication infrastructure, global visibility, and leadership in high-impact research. While Chinese papers are cited more often, the highest-impact publications are still largely concentrated in American institutions. Tools like GPT may further accelerate knowledge diffusion across regions, but asymmetries remain.

Lastly, we explored collaborations across disciplines, especially those connected to computer science (CS). After 2020, the US maintained stable CS-related interdisciplinary links. In contrast, China saw a notable increase—particularly in CS–Mathematics collaborations, which became much more prominent than other CS pairings. This suggests a targeted national push in strategically important areas. In the US, interdisciplinary CS work remains broadly distributed across many fields, often driven by the strengths of existing research communities. In China, policy incentives and the rapid growth of large language models likely played a role in promoting more focused and high-volume collaborations in key technical areas.

## References


[1] Bedoor AlShebli, Shahan Ali Memon, James A Evans, and Talal Rahwan. 2024. China and the US produce more impactful AI research when collaborating together. *Scientific Reports* 14, 1 (2024), 28576.

[2] Xiaojing Cai, Caroline V Fry, and Caroline S Wagner. 2021. International collaboration during the COVID-19 crisis: Autumn 2020 developments. *Scientometrics* 126, 4 (2021), 3683–3692.

[3] Caroline V Fry, Xiaojing Cai, Yi Zhang, and Caroline S Wagner. 2020. Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research. *PloS one* 15, 7 (2020), e0236307.

[4] Alberto Gómez-Espés, Michael Färber, and Adam Jatowt. 2024. Benefits of international collaboration in computer science: a case study of China, the European Union, and the United States. *Scientometrics* 129, 2 (2024), 1155–1171.

[5] Jenny J Lee and John P Haupt. 2020. Winners and losers in US-China scientific research collaborations. *Higher Education* 80, 1 (2020), 57–74.

[6] Loet Leydesdorff, Caroline S Wagner, and Lutz Bornmann. 2014. The European Union, China, and the United States in the top-1% and top-10% layers of most-frequently cited publications: Competition and collaborations. *Journal of Informetrics* 8, 3 (2014), 606–617.

[7] Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E* 69, 2 (2004), 026113.

[8] Jason Priem, Heather Piwowar, and Richard Orr. 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833* (2022).

[9] Xinyu Que, Fabio Checconi, Fabrizio Petrini, and John A Gunnels. 2015. Scalable community detection with the louvain algorithm. In *2015 IEEE international parallel and distributed processing symposium*. IEEE, 28–37.

[10] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. 2005. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 233–240.

[11] Jari Saramäki, Mikko Kivelä, Jukka-Pekka Onnela, Kimmo Kaski, and Janos Kertesz. 2007. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 75, 2 (2007), 027105.

[12] Li Tang. 2024. Halt the ongoing decoupling and reboot US-China scientific collaboration. *Journal of Informetrics* 18, 2 (2024), 101521.

[13] Junlong Zhang and Yu Luo. 2017. Degree centrality, betweenness centrality, and closeness centrality in social network. In *2017 2nd international conference on modelling, simulation and applied mathematics (MSAM2017)*. Atlantis press, 300–303.


## .1 Separated Degree Measures for Co-authorship Network

The degree centrality results show a shift in the collaboration structure between Chinese and U.S. institutions (See Fig.26, 27). Before 2020, their influence in the network was more balanced. After 2020, Chinese institutions became more central, while the role of U.S. institutions was slightly reduced. At the same time, the centrality values among top institutions became more even, suggesting the network is less dominated by a few major players. This could mean that collaborations are becoming more distributed and diverse across institutions.

The closeness centrality results show a similar pattern to degree centrality (See Fig.28, 29). Before 2020, Chinese and U.S. institutions were both very central in the network. U.S. affiliations like MIT and the Chinese Academy of Sciences had the highest scores. After 2020, Chinese institutions stayed strong, while the centrality of U.S. institutions became a bit lower.

At the same time, the closeness scores became more even across the top institutions. This suggests that the network is now more balanced, and not dominated by just a few nodes. It may mean that collaboration paths are more distributed, and more institutions are playing central roles in the global research network.

Also, the betweenness centrality results show that scores overall became lower after 2020 (See Figure.28, 29). This suggests that the academic collaboration network has become less centralized. In the earlier period, a few key institutions had a strong bridging role in connecting different parts of the network. After 2020, this role is more evenly shared across many institutions. It indicates that information flow and collaboration are now less dependent on a few central nodes and instead follow a more distributed structure.

By combing the results from all three centrality measures—degree, closeness, and betweenness—we observe a general trend toward a more balanced and distributed collaboration network after 2020. In all cases, the differences between top institutions became smaller, and more affiliations appeared to share similar levels of influence. This means that academic collaboration is no longer dominated by just a few central players. Among which, chinese institutions remained strong across all metrics, while the centrality of U.S. institutions showed a slight decrease.

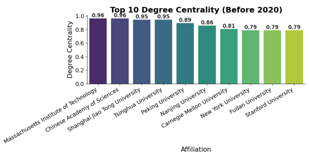


**Figure 26: Top 10 Affiliations for Collaboration Network by Degree Centrality (Before 2020)**

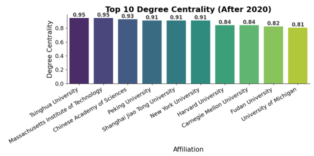


**Figure 27: Top 10 Affiliations for Collaboration Network by Degree Centrality (After 2020)**

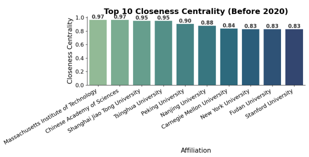


**Figure 28: Top 10 Affiliations for Collaboration Network by Closeness Centrality (Before 2020)**

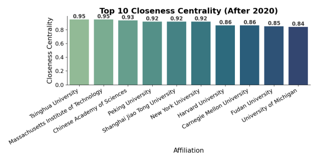


**Figure 29: Top 10 Affiliations for Collaboration Network by Closeness Centrality (After 2020)**

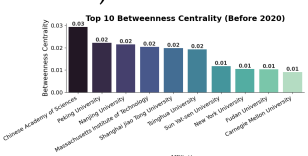


**Figure 30: Top 10 Affiliations for Collaboration Network by Betweeness Centrality (Before 2020)**

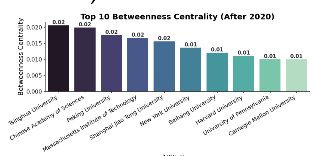


**Figure 31: Top 10 Affiliations for Collaboration Network by Betweeness Centrality (After 2020)**

## .2 Separated Degree Measures for Citation Network

See Fig.32, 33, these two figures show the changes in degree centrality in the citation network before and after 2020.

One obvious change is that Tsinghua University's degree centrality score went up a lot, from 1.24 to 1.53. This means it had more direct citation connections to other institutions—it became a more central node in the citation network. Another change is that more Chinese universities entered the top 10 after 2020. Before 2020, only a few Chinese institutions had high degree centrality. But after 2020, six out of the top ten were from China. This shows that Chinese research started to be cited more broadly, and not just focused on a few key institutions.

Overall, the degree centrality results suggest that citation patterns became more centered around Chinese affiliations after 2020.

The closeness centrality results reveal some clear trends in the citation network (See Fig.34, 35). Before 2020, both Chinese and U.S. institutions were present in the top ranks, with Tsinghua University and MIT leading the list. After 2020, the scores increased overall, especially for Chinese institutions. Tsinghua, for example, moved from 0.71 to 0.81, showing that it became even more central in the citation flow.

While U.S. universities like MIT and Stanford remained highly ranked, their relative closeness scores grew more slowly compared to their Chinese counterparts. This suggests that Chinese institutions have gained a stronger position in the citation network and are now more central in terms of information access and visibility.

At the same time, the gap between the top institutions and the rest widened slightly. This might point to a structure where a few key players—especially in China—are becoming more dominant in citation flows, while many other institutions are less connected.

Overall, the citation network seems to be shifting towards greater centrality for Chinese universities, with Tsinghua and the Chinese Academy of Sciences taking the lead. The influence of top U.S. institutions is still strong, but their centrality is not growing as fast.

Following the trends seen in degree and closeness centrality, the betweenness centrality results also show that Chinese institutions have become more central in the citation network after 2020. Tsinghua University, for example, had an exceptionally high score before 2020 (0.22), meaning it played a key role in connecting different parts of the network. After 2020, its score is still high, but the gap between institutions becomes smaller. U.S. institutions like Carnegie Mellon and MIT also keep their importance, but the overall scores are lower. This means that the citation network is no longer relying on just a few main hubs. Instead, the bridging role is now shared by more institutions, which points to a more balanced and distributed citation structure between China and the U.S.

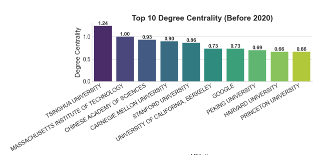


**Figure 32: Top 10 Affiliations for Citation Network by Degree Centrality (Before 2020)**

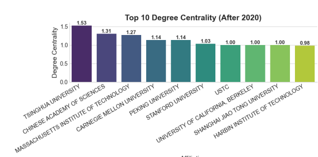


**Figure 33: Top 10 Affiliations for Citation Network by Degree Centrality (After 2020)**

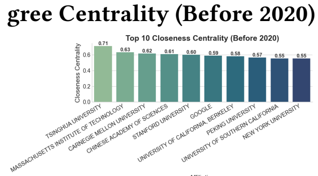


**Figure 34: Top 10 Affiliations for Citation Network by Closeness Centrality (Before 2020)**

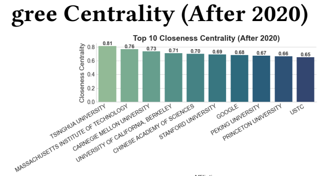


**Figure 35: Top 10 Affiliations for Citation Network by Closeness Centrality (After 2020)**

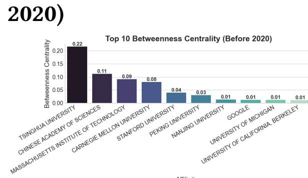


**Figure 36: Top 10 Affiliations for Citation Network by Betweeness Centrality (Before 2020)**

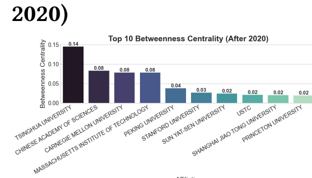


**Figure 37: Top 10 Affiliations for Citation Network by Betweeness Centrality (After 2020)**

**Table 4: Top 5 Co-occurrence Ratio Changes with Computer Science (Before 2020 vs. After 2020)**

| Country | Trend | Field | Change |
|---|---|---|---|
| 5*China | Decrease | Mathematics | −0.0892 |
| | Decrease | Philosophy | −0.0256 |
| | Decrease | Economics | −0.0247 |
| | Decrease | Biology | −0.0245 |
| | Decrease | Physics | −0.0220 |
| | Increase | Chemistry | +0.0182 |
| | Increase | Psychology | +0.0134 |
| | Increase | Engineering | +0.0130 |
| | Increase | Geology | +0.0061 |
| | Increase | Materials Science | +0.0045 |
| 5*US | Decrease | Mathematics | −0.0890 |
| | Decrease | Biology | −0.0311 |
| | Decrease | Philosophy | −0.0296 |
| | Decrease | Economics | −0.0193 |
| | Decrease | Political Science | −0.0079 |
| | Increase | Psychology | +0.0191 |
| | Increase | Physics | +0.0171 |
| | Increase | Medicine | +0.0160 |
| | Increase | Engineering | +0.0143 |
| | Increase | Geography | +0.0092 |