

## Study on the Information Exchange Structure of Selected Think Tank Websites –Based on URL Mutual Citation Network Analysis (Postprint)

**Authors:** Chen Wei, Li Xiuxia, Li Tingting

**Date:** 2017-11-05T00:00:00+00:00

### Abstract

[Purpose/Significance] URL citation represents a novel research methodology in webometrics that has emerged following the stagnation of link analysis development. Investigating URL mutual citation relationships among think tank websites enables the understanding of virtual proximity and distance between think tanks, as well as the exploration of roles and positions of different think tank websites. [Method/Process] This study employed Webometric Analyst 2.0 to acquire mutual citation data from 45 global think tank websites, conducting correlation analysis with think tank rankings to validate data reliability; constructed a URL mutual citation network among these 45 think tank websites, and utilized social network analysis to examine its primary characteristics and investigate network information exchange patterns among the think tanks. [Results/Conclusion] The findings reveal that three Chinese think tanks exhibit relatively limited network communication with other think tanks and lower efficiency. Finally, recommendations are proposed for the development of Chinese think tank websites.

### Full Text

#### Preamble

#### Analysis of the Information Communication Structure of Think Tank Websites Based on URL Cross-citation Networks

Chen Wei, Li Xiuxia, Li Tingting

School of Communication, Qufu Normal University, Rizhao 276826

## Abstract

**[Purpose/Significance]** URL citation represents a novel research methodology in webometrics developed in response to obstacles in link analysis. Examining URL cross-citation relationships among think tank websites can reveal their relative proximity in cyberspace and illuminate the roles and status of different think tanks. **[Method/Process]** We employed Webometric Analyst 2.0 to collect cross-citation data from 45 global think tank websites and validated data reliability through correlation analysis with think tank rankings. We then constructed a URL cross-citation network among these 45 think tank websites and applied social network analysis to reveal the network's principal characteristics and investigate information exchange patterns. **[Result/Conclusion]** The study demonstrates that three Chinese think tanks maintain relatively limited network communication with other think tanks, exhibiting low efficiency. Finally, we propose recommendations for developing China's think tank websites.

**Keywords:** think tank website; URL citation; social network; information communication

**Classification Number:** G350

## 1 Introduction

Think tanks refer to independent, relatively stable, and non-profit research institutions focused on national policy and public affairs. They provide scientifically sound recommendations for national strategy formulation and social development, constituting an essential component of modern national governance systems. Think tank websites serve as crucial external manifestations of think tank construction, functioning not only as primary channels for providing various information resources and services but also as important platforms for academic exchange among scholars. These websites do not exist in isolation; rather, they interconnect through hyperlinks, URL citations, title co-occurrences, and other mechanisms to form a social network that facilitates information exchange. Studying the information exchange structures among think tank websites can identify their relative proximity, uncover implicit information, and reveal deficiencies in think tank website development.

Most research on inter-institutional network relationships employs link analysis methods. However, as major commercial search engines have successively terminated complex link retrieval services, hyperlink data has become increasingly difficult to obtain, prompting many webometrics scholars to investigate URL citations and title co-occurrences. In 2011, Thelwall et al. first constructed URL citation networks and demonstrated that URL citation counts and institutional title mentions correlate with link counts, establishing them as viable alternatives for link network analysis [?]. Subsequent domestic scholars have explored URL citations theoretically and empirically. For instance, Chang Yanli examined URL co-occurrence networks among iSchool alliance members, finding that affiliations are influenced by geographical and linguistic factors [?]. Tang

Chuan et al. analyzed URL cross-citation networks among 100 Chinese university websites using social network analysis to explore network characteristics and influencing factors [?]. Wei Zhihao et al. conducted comparative research on China's 985 universities using URL and web title co-occurrence analysis, examining different characteristics represented by the two methods in various contexts [?]. In summary, current research on website URL citation networks primarily focuses on university and academic library websites, with relatively few studies addressing think tank websites.

This study collects URL citation data from 45 global think tank websites, conducts correlation analysis to validate the reliability of using URL citation data for analyzing inter-think tank relationships, and constructs a URL cross-citation network among these 45 think tank websites. Using social network analysis, we aim to reveal the network's main characteristics and investigate information exchange patterns among the 45 think tanks. Based on these findings, we conduct in-depth discussions to identify problems in Chinese think tank websites and propose targeted recommendations.

## 2.1 Selection of Research Subjects

This study examines think tank websites as research subjects, following principles of advancement and representativeness in sample selection. The Think Tanks and Civil Societies Program (TTCSP) at the University of Pennsylvania annually publishes the *Global Go To Think Tank Index Report*, widely regarded as the most authoritative report on global think tank development. This report provides comprehensive rankings based on the development status of think tanks worldwide. Therefore, we selected 45 think tank websites distributed across different countries and regions from the top 100 think tanks listed in the *2015 Global Go To Think Tank Index Report* [?] as our research sample, as shown in .

## 2.2 Data Acquisition and Processing

The URL citation frequency from website A to website B refers to the number of times website B's URL appears on website A. Manual retrieval and collection of URL cross-citation frequencies is both time-consuming and prone to accuracy issues. Therefore, this study utilizes Webometric Analyst 2.0 software for data collection. Developed by Professor Thelwall, Webometric Analyst 2.0 is an excellent URL data collection tool based on the Bing API [?].

Considering that URL cross-citation frequency magnitude reflects the corresponding institution's network information exchange activity, we did not standardize the collected URL cross-citation frequencies but directly constructed a  $45 \times 45$  matrix, where the data in row  $x$  and column  $y$  represents the citation frequency from think tank website  $x$  to think tank website  $y$ . Furthermore, this study only considers URL cross-citations between different websites, excluding self-citations; thus, all diagonal values in the matrix were assigned as 0.

### 2.3 Data Reliability Analysis

URL citation is a new method for exploring website relationships and remains in the experimental stage as an alternative to link analysis. Its validity must be demonstrated and analyzed. Validation of URL citation data primarily employs two approaches: first, comparing URL citation data with traditional link analysis data or other parameters; second, subjectively comparing URL citation data with actual circumstances [?]. Given current difficulties in obtaining link data, this study adopts the second approach. To make results more distinctive, we conducted correlation analysis between URL cross-citation data (specifically, in-degree rankings) and think tank rankings. Using SPSS for correlation analysis, the results are presented in .

The analysis reveals a significant and relatively strong correlation between think tank ranking and in-degree ranking, with a correlation coefficient of 0.668. This indicates that the collected URL cross-citation data effectively reflects the information resources and characteristics of the 45 think tanks and can be used for analyzing information exchange among think tank websites.

### 3 Construction and Analysis of Think Tank Website Cross-citation Network

Importing the think tank website cross-citation matrix into Netdraw generates the cross-citation network diagram shown in [Figure 1: see original paper]. In this network, nodes represent the 45 think tank websites, and edges represent URL citation relationships between them, with arrows pointing toward the cited party. Arrow direction indicates information inflow, while arrow origin indicates information outflow. Therefore, the URL cross-citation network also represents a think tank information flow and exchange network. Node size is determined by degree centrality, and edge thickness by citation frequency.

[Figure 1: see original paper] reveals that BK has the largest node, followed by CH and CEIP, indicating that these three think tank websites play pivotal roles in network information exchange, connecting various think tank websites, controlling information exchange directions, and serving as bridges in the network. Nodes for KAS, TI, ACPSS, RIE, and CH are slightly smaller than the three core think tanks but larger than others, suggesting they contribute significantly to information flow, particularly playing central roles in local information transmission within the network.

Regarding edge thickness, thick edges concentrate around core nodes, primarily within a group centered on BK, CEIP, and ACPSS, including KAS, TI, RIE, CH, PIIE, and ADBI. This indicates frequent information exchange among these websites in the URL cross-citation network. The thickest edge appears between ADBI and ACPSS. Further investigation reveals that ACPSS' s citations of ADBI mostly concern Asian affairs, suggesting ACPSS derives much of its information on Asian issues from ADBI. Notably, relatively thick edges also connect

peripheral network members such as AERC, NUPI, ACCORD, and SATTA with ACPSS, indicating substantial communication between these websites and ACPSS. This occurs because ACPSS, AERC, NUPI, ACCORD, and SATTA all originate from Africa; similar geographical environments, political and economic policies, and national cultures facilitate frequent exchanges among them. The diagram also shows that China's CIIS (China Institute of International Studies) performs prominently with relatively close network connections to other think tanks, while SIIS (Shanghai Institutes for International Studies) and CI-CIR (China Institutes of Contemporary International Relations) maintain fewer connections with other think tanks, occupying peripheral positions in the cross-citation network.

## 4 Analysis of Information Exchange Structure in Cross-citation Networks

To further explore each think tank website's network information exchange activity and its position in the network, and to reveal implicit network characteristics, this section applies social network analysis to examine the network structure. Since binary matrices often yield better results for certain social network analyses, we processed the original URL cross-citation matrix accordingly.

### 4.1 Overall Efficiency Analysis of Information Exchange in Cross-citation Networks

Analyzing the overall structural characteristics of social networks can reveal information transmission efficiency. This section examines network density and reachability. Network density measures the closeness of connections among network members [?]. Higher density indicates tighter connections, more frequent interactions, smoother information exchange, and higher efficiency. UCINET calculations show that the density of the think tank website cross-citation binary network is 0.1813. Meyer's research indicates that the maximum density value in actual network diagrams is 0.5 [?]. This suggests that while think tank websites maintain some connections, they are not particularly close, with relatively limited network information exchange and low efficiency.

Network reachability reveals network connectivity—the ease with which nodes connect with other nodes [?]. It is typically represented by the average distance between nodes. In highly reachable networks, shorter distances between nodes enable more efficient information transmission, facilitating information exchange among members. UCINET calculations show that the average distance in the think tank website URL cross-citation matrix is 2.223, indicating that communication between any two think tank websites requires passing through an average of three other websites. According to the conclusion that any two people are separated by no more than six intermediaries [?], this suggests moderate information transmission distance among think tank websites, allowing relatively accurate information dissemination.

## 4.2 Centrality Analysis of Information Exchange Structure

Centrality constitutes a crucial indicator in social network research. Centrality analysis can identify node importance, social prestige, and status, while revealing overall network decentralization or concentration patterns. Centrality indicators include point centrality and graph centralization; point centrality further divides into degree centrality, betweenness centrality, and closeness centrality [?]. Since this study focuses on each think tank website' s role in information exchange, we primarily analyze degree centrality and betweenness centrality in the cross-citation network.

### 4.2.1 Analysis of Information Exchange Based on Degree Centrality

Degree centrality refers to the number of nodes directly connected to a given node. Nodes with high degree centrality maintain close relationships with other nodes and possess greater network influence. The URL cross-citation network among think tank websites is a directed network, so degree centrality divides into out-degree centrality and in-degree centrality. In URL cross-citation networks, think tank websites with high out-degree centrality frequently cite other think tank websites and can absorb substantial network information; those with high in-degree centrality are frequently cited by other think tank websites, connect with numerous information sources, and can disseminate their own information outward. Nodes with both high in-degree and out-degree serve as knowledge flow hubs—widely receiving and sending knowledge, energizing the network and fostering innovative knowledge generation [?]. Calculations yield the degree centrality values for the 45 think tank websites, detailed in .

shows that BK has the maximum in-degree (31), followed by TI and CEIP, indicating that these think tanks' network information is widely cited by most think tank websites, serving as primary information sources. Moreover, BK and CEIP also exhibit large out-degree values, establishing them as authoritative websites controlling information flow throughout the network. Investigation reveals that both think tanks are American, globally renowned, long-established, deeply researched, and highly influential, with research scopes encompassing global security, economics, environmental and energy issues, and affairs across continents.

The analysis also reveals that KAS has the maximum out-degree (23), followed by ACPSS, SWP, and RIE, all with out-degree values exceeding 20. This indicates they absorb substantial network information and play important roles in information circulation. However, these four think tank websites exhibit relatively low in-degree, suggesting they cannot extensively disseminate information outward. Further investigation shows these four think tanks cite information from nearly all think tanks, covering political, economic, and cultural affairs across continents, yet are less frequently cited by others, with information dissemination limited to their own countries or nations with similar political and

economic environments. Therefore, these think tanks should enrich their website content while integrating information from other sources to attract more citations and enhance their status. For China's think tanks, CIIS (China Institute of International Studies) shows an out-degree of 15 and in-degree of 2, indicating good information absorption but weak outward information transmission. Its website information is not widely cited or disseminated, possibly due to insufficient innovation, outdated content, language barriers, and other factors. Meanwhile, CICIR (China Institutes of Contemporary International Relations) and SIIS (Shanghai Institutes for International Studies) exhibit small out-degree and in-degree values, demonstrating weak performance in both network information absorption and dissemination.

#### 4.2.2 Analysis of Information Exchange Based on Betweenness Centrality

Betweenness centrality characterizes a node's capacity to lie on the shortest paths between other nodes. A node (actor) with higher betweenness centrality can control more resources and possesses greater interpersonal influence [?]. UCINET calculations yield the betweenness centrality values for each think tank website, shown in . The results indicate that BK, CEIP, IISS, and KAS—think tank websites from Europe and America—exhibit high betweenness centrality values exceeding 200, positioning them on the shortest paths of network information exchange among multiple think tank websites. These think tanks possess strong control over network information resources and serve as intermediaries in overall network information exchange.

For China's three think tanks—CIIS (China Institute of International Studies), CICIR (China Institutes of Contemporary International Relations), and SIIS (Shanghai Institutes for International Studies)—betweenness centrality values are small, indicating weak information control and limited influence for Chinese think tank websites in the URL cross-citation network.

### 4.3 Analysis of Small Group Communication in Cross-citation Networks

In cross-citation networks, small groups refer to clusters with frequent mutual citations, active knowledge exchange, and highly correlated research directions [?]. Analyzing these small groups can clarify the internal structure of cross-citation networks and each group's role and status. This section sequentially conducts component analysis, clique analysis, clique overlap analysis, and block model analysis on the think tank website URL cross-citation network.

#### 4.3.1 Component Analysis

Component analysis identifies network components. In directed networks, components can be weak (ignoring relationship direction) or strong (considering

direction) [?]. This section conducts strong component analysis on the think tank URL cross-citation network. Results show that the 45 network members divide into four components: the largest component contains 42 nodes, while AERC, ATL, and CEDICE each form separate components. CEDICE exhibits zero in-degree and out-degree, representing an isolated point; AERC has zero out-degree, and ATL has zero in-degree, both occupying peripheral network positions. Since component analysis does not reveal detailed internal group structures, we proceed to clique analysis.

### 4.3.2 Clique Analysis

Unlike components, cliques require that all points within a clique be mutually adjacent and not contained within a larger clique [?]. Before conducting clique analysis, we first symmetrize the directed data. This study employs UCINET's "Symmetrizing Method→Minimum" to symmetrize the directed matrix using reciprocal relationships. Clique analysis on the symmetrized matrix (with minimum membership set at 3) yields 17 cliques.

### 4.3.3 Clique Overlap Analysis

Analysis of clique results reveals substantial overlap among cliques. For instance, BK appears in 14 cliques, indicating that think tank BK communicates with most cliques and dominates network information exchange. Think tank CEIP appears in 10 cliques, demonstrating close connections with most clique members. Notably, think tank websites including KIEP, CIIS, KDI, and YDI do not belong to any clique, identifying them as isolates in the network. Investigation shows these think tank websites primarily originate from non-English-speaking countries, suggesting language barriers may prevent their inclusion in any clique. Therefore, these non-English-speaking think tank websites should prioritize developing English versions to ensure smooth communication.

### 4.3.4 Block Model Analysis

Clique overlap may obscure underlying structures [?]. To further extract internal network structural information, we conduct block model analysis—a method for studying network position models and describing social roles algebraically [?]. Using UCINET's CONCOR program to analyze the think tank URL cross-citation matrix yields the block diagram and block density matrix shown in [Figure 2: see original paper] and , respectively. [Figure 2: see original paper] reveals that the think tank URL cross-citation network can be divided into eight blocks.

With the overall network density of 0.1813 as the critical value, we compare each block's density in to this threshold, converting the block density matrix into an image matrix. The simplified diagram of this image matrix is presented in [Figure 3: see original paper]. The diagram shows that connections among

the eight subgroups formed by the 45 think tank websites are not close. Subgroups 1 and 7 serve as information sources: the former outputs information to subgroups 2 and 5, while the latter transmits information to subgroups 4 and 5. Subgroup 1 contains the most think tanks. Comprehensive analysis of its characteristics reveals that most think tanks in block 1 belong to Europe and America, with broad research scopes covering global economics, livelihood issues, health, education, and other topics. Examples include BK, CH, CI, and CEIP, which rank within the top 20 across multiple think tank categories and demonstrate strong comprehensive strength [?]. Subgroup 2 can be considered network brokers, both transmitting and receiving information. Analysis shows that think tanks in subgroup 2 primarily research domestic and internationally relevant issues, with relatively narrow research scopes. For example, KAS mainly studies various EU affairs and transatlantic development and cooperation, aligning with subgroup 2' s broker positioning by absorbing information from subgroup 1, integrating it, and transmitting it to subgroup 5. Subgroup 5 is a receptive subgroup, accepting information from subgroups 1, 2, 3, and 7. Subgroup 5 includes only ACPSS and SIIS, from Egypt and China—both developing major countries. Consistent with these nations' actual development patterns of absorbing external scientific information, ACPSS and SIIS primarily absorb various external information. Therefore, these two think tanks need to focus on innovation, continuously produce new research outcomes, and promote their own development and national soft power while absorbing network information from other think tanks.

Further analysis of the think tank URL cross-citation network' s small groups and website content suggests that think tank network information exchange may also be influenced by website language and think tank scale. To explore these relationships, we conducted correlation analysis between URL cross-citation network in-degree indicators and website language/think tank scale indicators. The language indicator is ordinal: level 1 for English-language websites, level 2 for non-English websites with English pages, and level 3 for non-English websites without English pages. Think tank scale is represented by employee count. Investigation and statistics for each think tank website yielded language status and employee numbers. Prior to correlation analysis, we examined the statistical properties of both datasets, which exhibited skewed distributions; therefore, we calculated Spearman correlation coefficients using SPSS. Results are presented in . The analysis shows that think tank language status significantly correlates with in-degree ranking, while employee count does not significantly correlate with in-degree ranking. This indicates that website language constitutes a factor influencing think tank network information exchange, with English-language think tanks engaging in more network communication, whereas think tank scale has minimal impact on network information exchange.

## 5 Conclusions and Implications

This study utilized Webometric Analyst 2.0 software to collect URL cross-citation data from 45 global think tanks, validated this new webometric data source, and explored global think tank network information exchange from a novel perspective, effectively addressing the challenge of analyzing think tank website relationships amid scarce link data. The following conclusions emerge:

- (1) As an alternative and supplement to link data, URL data can be used for webometric analysis to a certain extent. Through correlation analysis, we found that URL data rankings significantly correlate with think tank rankings, confirming that our collected URL data can serve as valid analytical data for inter-think tank network information exchange.
- (2) The think tank URL cross-citation network exhibits low density, indicating limited network information exchange and low efficiency. However, information transmission distance among think tank websites is moderate, enabling relatively accurate information dissemination.
- (3) In the think tank URL cross-citation network, think tank websites serving as information sources and brokers predominantly originate from Europe and America, indicating that these think tanks control network resources and represent network authorities. In contrast, think tanks from other continents, particularly the three Chinese think tanks, primarily absorb network information while scarcely outputting information, resulting in limited influence in the URL cross-citation network.
- (4) Through small group analysis and website content examination of the think tank URL cross-citation network, we find that think tank information exchange is influenced by language and research scope, with English-language think tanks conducting extensive research and engaging in more communication. However, think tank scale has minimal impact on network information exchange.

Based on these conclusions and think tank website content analysis, we derive the following implications:

URL citations among think tank websites enable information dissemination and facilitate network information and knowledge exchange. However, the three Chinese think tanks exhibit poor network communication performance, indicating limited influence of Chinese think tank websites on the Internet.

First, we must enhance Chinese think tank strength and expand academic research content. Investigation of core influential think tanks reveals that their research content is extensive, covering not only macro-level issues such as national politics and military affairs but also micro-level issues closely related to people's lives, including livelihood and healthcare. Chinese think tank research content is relatively narrow, focusing primarily on national policy formulation with minimal attention to details of public life and insufficient alignment with

contemporary development. Therefore, Chinese think tanks should expand their research content to achieve multi-faceted, multi-level development.

Additionally, we must strengthen Chinese think tank website construction, enrich content, and particularly emphasize English website development to ensure smooth communication with other websites. Investigation of core think tank websites such as BK reveals that these sites embed social tools like blogs and communication platforms to facilitate research dissemination. In this era of widespread social tool application, this approach merits study and emulation.

## References

- [?] Thelwall M, Sud P. A Comparison of Methods for Collecting Webcitation Data for Academic Organizations[J]. *Journal of the American Society for Information Science and Technology*, 2012(63):1960-1972.
- [?] Chang Yanli. iSchools Alliance College URL Co-citation Network Link Analysis[J]. *Journal of Intelligence*, 2013, 32(4):108-112.
- [?] Tang Chuan, Liu Chunjiang, Xu Jing. Research on Structural Characteristics and Related Factors of Chinese University Website URL Citation Networks[J]. *Journal of Intelligence*, 2014,33(4):46-51.
- [?] Wei Zhihao, Shi Guoliang. Comparative Empirical Study on Characteristics of URL and Web Title Co-occurrence Data: Taking “985 Universities” as an Example[J]. *Library Science Research*, 2015(23):80-85.
- [?] TTCSP. 2015 Global Go To Think Tank Index Report[EB/OL].[2016-2-9]. [http://repository.upenn.edu/cgi/viewcontent.cgi?article=1009&context=think\\_{tanks}](http://repository.upenn.edu/cgi/viewcontent.cgi?article=1009&context=think_{tanks}).
- [?] Thelwall M, Pardeep S. Webometric Research with the Bing-Search API 2.0[J]. *Journal of Informetrics*, 2012,6(1):44-52.
- [?] Deng Sanhong, Wang Hao. Analysis of 985 University Name Co-occurrence and Referral Relationships Based on Title Mention[J]. *Journal of the China Society for Scientific and Technical Information*, 2014(12):1322-1334.
- [?] Stanley Wasserman, Katherine Faust. *Social Network Analysis: Methods and Application*(M). London: Cambridge University Press, 1994.
- [?] Liu Jun. *Introduction to Social Network Analysis*[M]. Beijing: Social Sciences Academic Press, 2004:114-116,62,114-116.
- [?] Yue Zenghui, Fang Shu. Research on University Library Co-link Networks Based on SNA[J]. *Information and Documentation Services*, 2012(6):61-65.
- [?] Stanley Milgram. The small world problem[J]. *Psychology Today*, 1967,(2):60-67.
- [?] Wang Xiaohong, Zhang Baosheng. Analysis of the Effect of Knowledge Network Structural Characteristics on Knowledge Flow[J]. *Value Engineering*, 2010,29(2):11-13.
- [?] Liu Jun. *Lecture Notes on Whole Network Analysis: A Practical Guide to UCINET Software*[M]. Shanghai: Shanghai People' s Publishing House, 2009:100,125,150-153.
- [?] Zhang Qin, Ma Feicheng. Research on Knowledge Exchange Structure in Domestic Knowledge Management Field: Analysis Based on Core Author

Cross-citation Networks[J]. Journal of the China Society for Scientific and Technical Information, 2012,31(9):925-933.

[?] Guo Jinlong, Xu Xin. Social Network Analysis of Domain Blogs: Empirical Study Based on Library and Information Science and Internet Blogs[J]. Library and Information Service Online, 2012(1):1-8.

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv –Machine translation. Verify with original.*