ISSI 2023 Proceedings, v1, pp 345–362
DOI: 10.5281/zenodo.8305883 https://cns-iu.github.io/workshops/2023-07-02_issi/

Understanding and identification: The life cycle stages of research team development

Ruinan Li[1], Raf Guns[2], Tim C. E. Engels[3], Lin Zhang[4], Ying Huang[5]

[1]*liruinan93@126.com*
Center for Studies of Information Resources, School of Information Management, Wuhan University, Wuhan 430072, China
Center for Science, Technology & Education Assessment (CSTEA), Wuhan University 430072, China
Center for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp (Belgium)

[2]*raf.guns@uantwerpen.be,* [3]*tim.engels@uantwerpen.be*
Center for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, (Belgium)

[4]*zhanglin_1117@126.com,* [5]*ying.huang@whu.edu.cn*
Center for Studies of Information Resources, School of Information Management, Wuhan University, Wuhan 430072, China
Center for Science, Technology & Education Assessment (CSTEA), Wuhan University 430072, China
Center for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Leuven 3000, Belgium

Abstract
This study proposes a new framework for understanding and identifying the life cycle stages of research teams by studying the evolution of team effectiveness. We first develop a conceptual model to describe a basic 'life cycle' of research teams, and then test the model for the case of the NSFC Innovative Research Groups, to illustrate the strengths and potential of this method. Through the framework, we can observe the overall growth curve of team effectiveness and how it has evolved over time in the different life cycle stages.

Keywords: Research team; Life cycle; Team effectiveness; Generalized principal component analysis (GPCA); Change point detection

Introduction
Scientific collaboration is becoming increasingly widespread and intensive with the need for scientific and technological development. Research teams formed by close scientific collaborations have played the primary force role in scientific research and are profoundly influencing knowledge production and the advancement of science and technology. Especially in contemporary team science, scholars are more curious about the temporality of teams: how teams evolve and mature, and how team dynamics play out over time (Shuffler, Salas, & Rosen, 2020). Research teams or groups are the smallest organizational units of the science system, and studying the development of research teams is of great value to enhance our understanding of the functioning and evolution laws of the science system (Braam & Van den Besselaar, 2010). Also, it can be helpful to policy development for science and technology management.

According to the previous literature, team changes and development tend to follow a similar life cycle, consisting of different phases. Teams at a given development stage tend to exhibit a common pattern of actions and behaviors related to team tasks and interpersonal relationships, and a similar level of team effectiveness (Peralta, Lourenço, Lopes, Baptista, & Pais, 2018). For instance, teams functioning at the higher stages of development tend to be more productive and have healthier and more satisfied members (Peralta et al., 2018). Many theoretical models and frameworks that describe how teams evolve through this process have been proposed. It can be traced back to Tuckman's stages of group development (Tuckman, 1965). In 1965, psychologist Bruce Tuckman developed a model of group development and thought that four general stages of group development could be named forming, storming, norming, and performing (Hurt & Trombley, 2007; Tuckman, 1965). In 1977, Tuckman and Jensen (1977) updated this model and added a final stage, adjourning. Therefore, the sequence, theoretically the same for every group, consists of forming, storming, norming, performing, and adjourning (Gersick, 1988). As with other kinds of teams, research teams follow certain patterns in their life cycle stages of development. For instance, Braam and Van den Besselaar (2010, 2014) have shown that the growth of activities at the research group level follows an S-curve pattern by analyzing the long-term development of research groups and institutes.

To make a judgment on the life cycle stages of a research team, scholars often consider multiple dimensions, such as the trust between team members, social cohesion, and relationship conflict, and through the questionnaire to get the answer (Peralta et al., 2018). This method is advantageous but also difficult and labor-intensive to quantify because one needs to investigate all team members and situations. Therefore, others limit themselves to using quantitative indicators to judge the stages of the team life cycle. For example, Robert Braam and Peter van den Besselaar developed a lifecycle-based theoretical model for analyzing the long-term development of research groups and institutes, and proposed three bibliometric indicators to describe the research dynamics of the case institute, including growth, activity profile stability, and focus (Braam & Van den Besselaar, 2010, 2014). The main advantage is that this method is easy-to-use. However, there are two weaknesses to this method. On the one hand, based on the analysis results, we only know that the team development process can be divided into several phases (e.g., the first, second, third, fourth phase, and so on.), but we don't know which one lifecycle stage these phases should belong to, such as the initiation, growth, maturity, and saturation stage. On the other hand, Braam and Van den Besselaar only used the number of team outputs to determine stages of team development. In addition to the familiar types of articles and reviews, they also include book chapters, lectures, reports and presentations, and so on. It is a strong point that scholars considered different kinds of outputs. However, it may be doubtful whether these can be added directly in calculating team outputs. Perhaps it makes more sense to consider them separately and/or use some form of weighting. Team effectiveness and team performance are comprehensive concepts, and they involve aspects beyond raw productivity. So, if we use one or more comprehensive indicators to observe the development of research teams and detect their life cycle stages, it will be valuable and useful.

An effective team is one that produces high performance, high team member satisfaction, and team viability (Towler, 2020). The committee on the Science of Team Science (National Research Council, 2015) defines team effectiveness, also referred to as team performance, is a team's capacity to achieve its goals and objectives. This capacity to achieve goals and objectives leads to improved outcomes for the team members (e.g., team member satisfaction and willingness to remain together) as well as outcomes produced or influenced by the team. Guzzo and Dickson (1996) also thought that team effectiveness is indicated by (a) team-produced outputs (quantity or quality, speed, customer satisfaction, and so on), (b) the consequences a team has for its members, or (c) the enhancement of a team's capability to perform effectively in the future. According to these definitions, we can know that team performance is a component of team effectiveness because teams are generally considered to "perform well" when they yield superior outputs (Towler, 2020). Additionally, it also involves the collaboration dynamics within teams. Thus, some scholars also believe that team effectiveness can be divided into two types: team outcome effectiveness and team process effectiveness (Collins & Parker, 2010). In this study, we will concentrate mostly on team outcome effectiveness and team collaborative relationships due to the difficulty of obtaining all team process data. And we want to present a long-term collaboration among team members through co-authorship networks. Overall, we are trying to present team effectiveness in terms of three dimensions: productivity, impact, and collaboration. It is easy to understand that during the development and evolution of research teams, their team effectiveness is changing in response to internal and external conditions. In this study, our aim is to analyze research teams' development and team effectiveness over time based on the team publications and identify their life cycle stages. This may help decision-makers and managers estimate its future development trends.

Framework and methodology

In this paper, we examine how the team effectiveness of research teams changes over the lifecycle development of teams. The research framework is shown in Figure 1.

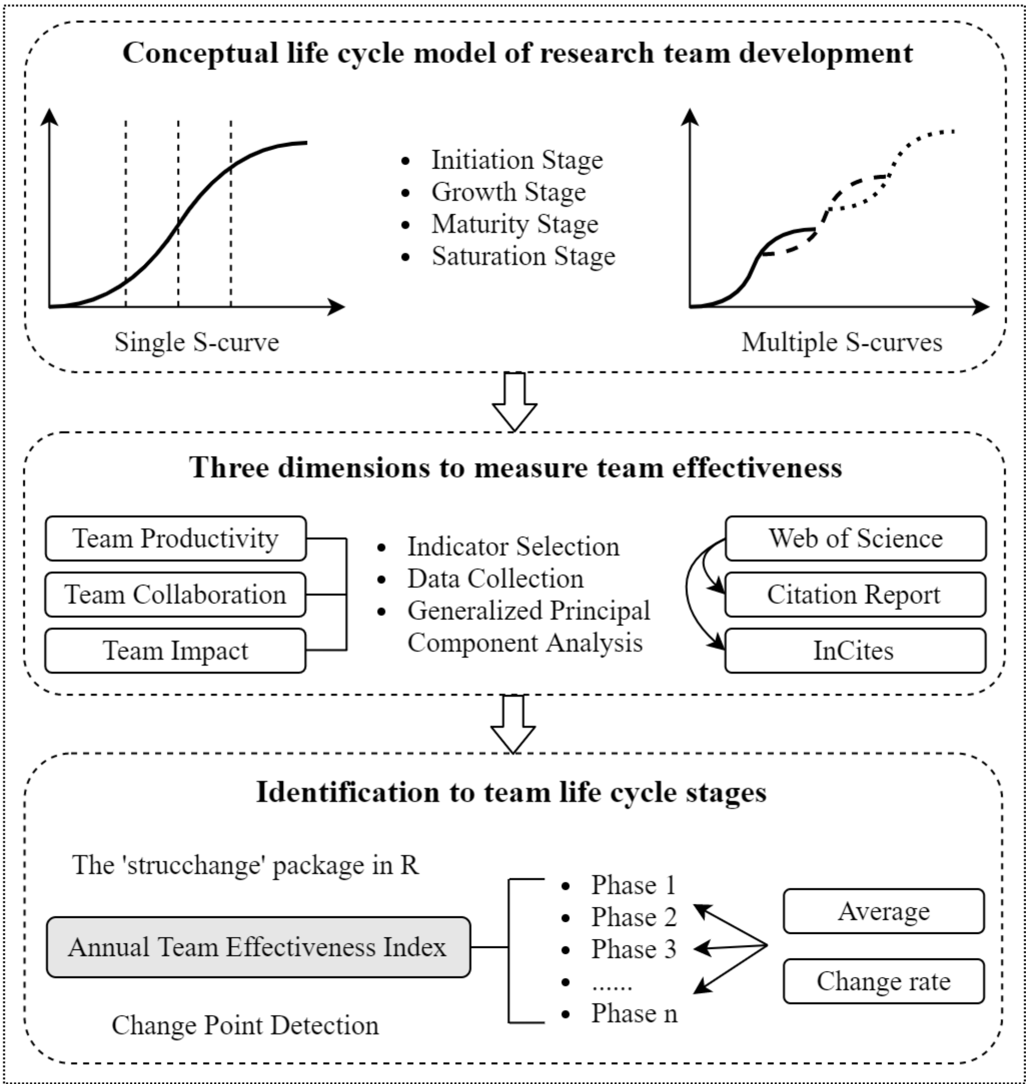

Figure 1. The research framework

Understanding research team development: Conceptual life cycle model

Every organizational research unit faces internal and external conditions that limit its ability to achieve goals. Because resources are generally limited, the size of a team's activities will grow to a certain limit defined by the maximum amount of resources available (Braam & Van den Besselaar, 2010). The growth pattern of activities in such cases will have the shape of an S-curve (Braam & Van den Besselaar, 2010; Price, 1963). Therefore, there is one assumption before the conceptual model of team life cycle development is proposed. That is, the research teams develop in stable conditions with limited resources, and team development follows a pattern of moving from lower to higher stages. For example, there are no policies that have a significant influence on team development from one phase to another one. In addition, the reason why there is a limit to the growth of team effectiveness is that academic resources are limited in each research area, and every scholar has limited energy. Especially, the limited career age of the most senior team members can play a role in the team dynamics. Thus, in the process of team development, the level of team effectiveness is initially low, and the first growth is slow, followed by rapid growth, and then followed by slow growth until a plateau is reached.

According to the above description, we know that the ideal growth of team effectiveness will also follow an S-curve. This point has also been supported in relevant studies (Braam & Van den Besselaar, 2010, 2014; Tuckman, 1965; Tuckman & Jensen, 1977). S-curves have four stages, initial slow growth (initiation/birth), rapid growth (acceleration/growth), late-stage slow growth (deceleration/maturing), and no growth (decline). Figure 2 shows the hypothesized basic evolutionary patterns of team effectiveness over time. In this case, we can divide the long-term development of research teams into four phases: initiation, growth, maturity, and decline (see Figure 2). When the team effectiveness of research teams follows this development pattern, we can judge the life cycle stage of a team based on the absolute value and change rate of team effectiveness. And the specific judgment rules are shown in Table 1.

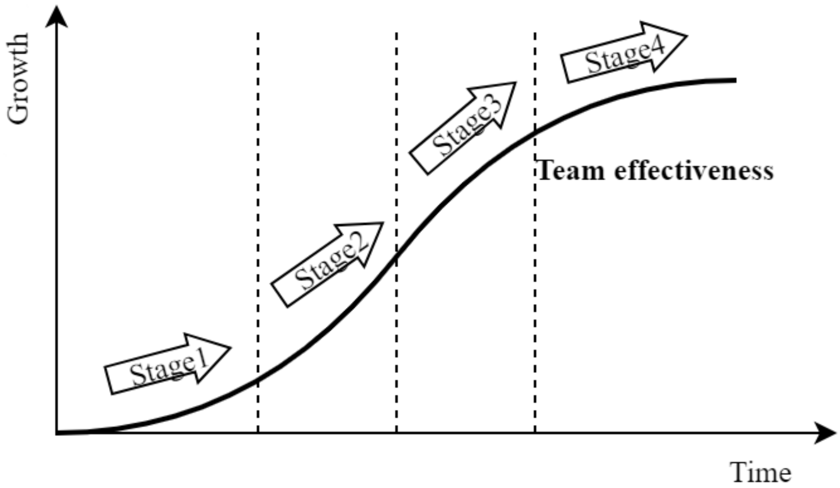

Figure 2. Conceptual life cycle model of research teams based on team effectiveness

However, the stages of team development in the above conceptual models are ideal representations of teams' life cycles. Actually, since the team cannot live in ideal conditions, teams do not necessarily have all lifecycle stages, and there may even be an escalation. For example, it is likely to have double-peaking or muti-peaking trends in the growth curve of team effectiveness (as shown in Figure 3).

Table 1. Rules for determining the life cycle stages of research teams under ideal conditions

Team life cycle stages	*The evolutionary characteristics of team effectiveness*	
	Trend of change	*Rate of change*
Initiation	Growing from zero or near zero	Positive numbers close to zero
Growth	Upward trend	Max.

Maturity	Fluctuating up and down (Stable trend)	Lower than that in the growth stage
Decline	Stop the rising or downward trend	Near zero or negative numbers

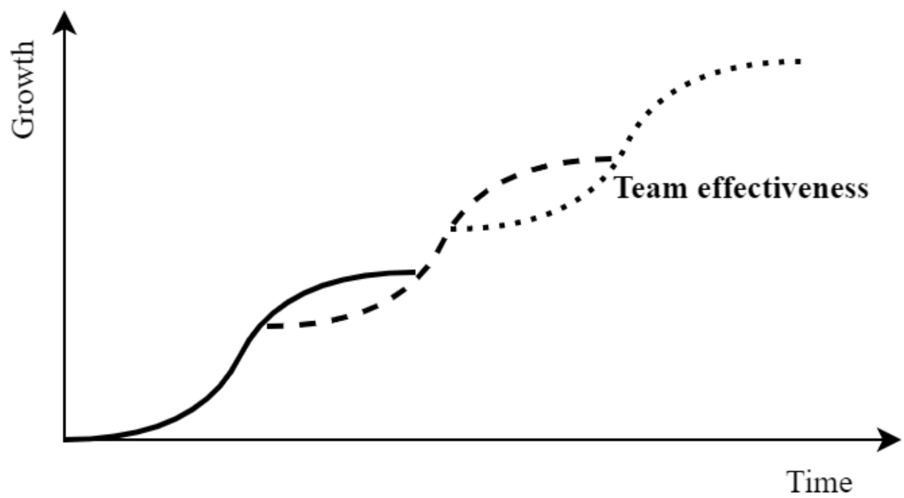

Figure 3. Conceptual life cycle model of multiple S-curves for research teams

Measurement of team effectiveness

For the evaluation of the research activity, bibliometric indicators are always widely used (Curry et al., 2022; Kulikowski, Przytuła, & Sułkowski, 2022). To measure research performance, we often use several bibliometric indicators from two aspects: productivity and impact (number and quality) (Nane & Bachasingh, 2019). For productivity, it is almost well established that the publications produced by a research team over time can represent the team's output capacity. In terms of the impact, it often only involves the scientific impact. And citations are widely seen as an indicator of the impact, significance, and influence of academic publications (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Garfield, 1955; Nane & Bachasingh, 2019). Since a single indicator only measures one aspect of team performance, we should try to use a range of different indicators to provide an adequate measure of team performance. The indicators for performance assessment can be sorted into different categories: productivity and impact, comparative and normalized (percentiles, normalized citation impact, influence) (Potter, 2016). In this study, for a more meaningful and comparable analysis, both the productivity and impact indicators (such as the number of publications and citations) and also normalized values (such as the Category Normalized Citation Impact) should be considered. For the collaborative relationships between team members, we will use several network indicators to measure the degree of team collaboration, such as network density and clustering coefficient. Among them, network density can be used to measure the degree of cohesion among team members. The clustering coefficient can be used to determine the extent to which team members are clustered together. The following indicators will be used to measure team effectiveness in this study, and it is as shown in Table 2 (Clarivate, 2021; Franceschet, 2009).

Table 2. Dimensions and indicators for measuring team effectiveness

Dimensions	*Indicators*	*Meaning of indicators*	*Types*
Team productivity	(1) Web of Science Documents	The total number of Web of Science Core Collection papers for that entity. This count includes all document types, and it is a measure of productivity.	Productivity Indicators
Team impact	(2) Times Cited	The number of times a set of Web of Science Documents has been cited. This indicates the total influence of a set of publications.	
	(3) Citation Impact	Calculate the citation impact of a set of documents by dividing the total number of citations by the total number of publications. Citation impact shows the average number of citations a document received.	Impact Indicators
	(4) Cumulative Citations per Year	This is a different version of Times Cited per year, depicting the number of citations received each year. In other words, the number of documents from each year that cite (reference) a paper published by a selected institution in the same year or prior. For example, for any source document, count citing documents, and group by the publication year of the citing document.	
	(5) H-index	A team has an h-index if it has at least h publications for which it received at least h citations.	
	(6) Category Normalized Citation Impact (CNCI)	The CNCI of a document is calculated by dividing the actual count of citing items by the expected citation rate for documents with the same document type, year of publication, and subject area. The CNCI of a set of documents, for example, the collected works of a research team, is the average of the CNCI values for all the documents in the set.	Normalized Indicators
Team collaboration	(7) Network Density	Network density refers to the proportion of connections in a network compared to the total number of possible connections.	Network indicators
	(8) Clustering Coefficient	The clustering coefficient of a node is defined as the fraction of pairs of the node's neighbors that are connected to each other, divided by the total number of possible pairs of neighbors. In other words, it measures how likely the neighbors of a node are to be connected to each other. The clustering coefficient of a network is then the average of the clustering coefficients of all its nodes.	

Only if the research teams are engaged in the field of basic research can these indicators relevant to publications be used to gauge the effectiveness or performance of the teams because basic researchers typically publish their research findings in papers rather than in patents or other forms. When our samples satisfy this requirement, we use the names and affiliations of team members to search the team's publications in the Web of Science database. After the data cleaning to author names, we obtain the updated team publication data. Based on the publication data, we then create citation reports that include information about team members' publications and the corresponding annual citations. Additionally, in order to obtain the values of normalized indicators, we saved the search retrieval to the InCites database and downloaded the indicators data for each of the team's publications. And then, we create team collaborative networks based on the team publications and calculate the values of network indicators. Finally, we calculate and save the values of all indicators in each year for all teams. These data and results will be the basis for data analysis in this study.

After obtaining the above data, we first need to calculate the specific values of team effectiveness for different years and draw their curves over time. For the measurement of team effectiveness, we should use multiple indicators to character the multiple dimensions of team performance and team collaboration according to the above description. While it is wise to use a variety of metrics, it is unsystematic and confusing to have too many of them, in particular when metrics are highly correlated (Franceschet, 2009). In this study, these bibliometric indicators are calculated based on the number of team publications and citations. They are not completely independent, but are correlated with each other, which also has been discussed and verified in other literature (Franceschet, 2009; Nane & Bachasingh, 2019). Therefore, to avoid information redundancy in the results, we should consolidate multiple indicators to construct a weighted indicator to measure team effectiveness. Principal component analysis (PCA) is a multivariate statistical technique used to reduce a multidimensional space to a lower dimension (Franceschet, 2009; Giordani, 2018; Wikipedia, 2023). This technique is particularly useful in processing data where multi-collinearity exists between the variables. Because it can increase the interpretability of data while preserving the maximum amount of information. In this study, we will apply the PCA to reduce the dimensionality of our dataset and build a new comprehensive index to measure team effectiveness. Based on this new index, we can then calculate the annual team effectiveness index for each team.

Identification of team life cycle stages

Since the growth of team effectiveness tends to possess a stage-specific characteristic in the team development process, we can judge the change points between the different stages based on the level and change rate of team effectiveness. There are many change point detection tools to identify time steps when some statistical property of the time series changes, such as the changes in the mean value, standard deviation, or slope (linear trend) of continuous variables, as well as changes in the mean of count variables. In our study, we need to detect those change points that divides each time series into segments, where the values within each segment have a similar linear trend (slope and intercept). The ‘strucchange’ package in R can be used to test, monitor, and date structural changes in (linear) regression models. And we will apply this tool to detect the change points in the slope (or linear trend) of team effectiveness.

They are also the points to divide the different stages of the team life cycle. Based on these change points between multiple stages, we can further calculate the average and change rate of team effectiveness in each stage during the life cycle development of research teams. While we consider the average and change rate of team effectiveness at each stage, we also need to pay attention to the sequence in which the different stages occur. For example, the identification and naming of the different stages should follow the sequence from initiation - growth - maturity - decline stage. Finally, we can determine which team life cycle stage each of those stages detected should belong to according to the rules for judging the life cycle stages of research teams under ideal conditions in Table 1.

Case study: NSFC Innovative Research Groups

Samples and data

According to Rey-Rocha et al. (2006), the term 'team' in a scientific context may be defined from two perspectives: in output-based studies, a team is defined based on co-authorship, while in input-based studies, a team is based on existing administrative arrangements. In general, these two definitions of team correspond to virtual and physical teams, such as laboratory or project teams. In this study, we prefer virtual teams because we want to observe the evolution of research teams and their life cycle development, and collaborative networks can easily present long-term co-authorship relationships between team members. To identify research teams accurately and obtain enough samples, we think it is feasible to identify team members who have had a real and stable collaboration relationship with one another through project teams and consider them as virtual teams. In a networked environment, these team members gradually form a large collaborative network through collaborative publications. This is in line with the definition of virtual teams. The following is a brief description of the project teams of our samples.

The innovative research group projects of the National Natural Science Foundation of China (NSFC) mainly support outstanding young and middle-aged scientists as academic leaders and research backbones who jointly carry out innovative research around one significant research direction (Wu, 2019). The fund's purpose is to build a team towards the forefront of the international scientific community. In terms of samples, we choose 111 research teams built on the projects which were approved from 2007 to 2009 as the samples for the final selection. In terms of data, we use the collaboration publications by core project members as the team outputs. Finally, we create citation reports in the Web of Science and download the indicator data in the InCites database for the team's publications. This will be the foundation for data analysis in this study. Table 3 describes our data for the sample of 111 research teams. The number of team members on each project team is referred to as the team size. Team publications are the collaborative publications that team members publish in the current institutions. As shown in Table 3, these 111 research teams involve 1065 scholars and 31744 publications in total. They are spread over eight research fields. An individual team produced at least two publications per year and at least fifty publications over the course of the team's existence.

Annual Team Effectiveness Index

Calculate annual values of indicators

Based on the above data of the research team, we can calculate the annual values of indicators for each team. The results are shown in Table 4, and the density distribution curves of indicators are present in Figure 4, and the area under these curves is 1. From the density plot, we can know that these samples' indicator values are primarily distributed in front of the x-axis, indicating that most of the indicator values are at lower levels and that these teams have experienced lower levels for a longer period of time.

Table 3. A basic description of the data from 111 research teams

Funded fields		*Team size*	*Team publications*	*Annual average publications*	*Team Duration*	*Number of teams*
	Max.	12.00	1663.00	55.43	36.00	16
	Min.	7.00	62.00	2.33	16.00	

A. Mathematical and Physical Sciences	Mean	9.69	324.69	12.56	24.44	
	Median	10.00	110.50	6.09	23.50	
B. Chemical Science	Max.	11.00	1010.00	33.67	32.00	20
	Min.	6.00	56.00	2.33	19.00	
	Mean	9.35	338.85	12.75	25.90	
	Median	10.00	263.50	9.50	27.00	
C. Life Sciences	Max.	11.00	305.00	15.25	31.00	14
	Min.	4.00	54.00	2.48	17.00	
	Mean	8.93	126.29	5.65	23.21	
	Median	10.00	101.50	3.90	23.00	
D. Earth Sciences	Max.	10.00	332.00	13.28	30.00	16
	Min.	9.00	61.00	2.44	17.00	
	Mean	9.88	173.19	7.19	24.06	
	Median	10.00	157.00	6.78	24.00	
E. Engineering and Materials Science	Max.	14.00	1174.00	43.48	32.00	24
	Min.	7.00	53.00	2.52	15.00	
	Mean	10.08	394.17	15.31	24.71	
	Median	10.00	360.00	14.81	27.00	
F. Information Sciences	Max.	13.00	877.00	32.48	32.00	16
	Min.	6.00	51.00	2.55	16.00	
	Mean	9.63	320.94	12.12	25.25	
	Median	10.00	190.50	8.23	26.00	
G. Management sciences	Max.	10.00	106.00	7.07	21.00	2
	Min.	10.00	75.00	3.57	15.00	
	Mean	10.00	90.50	5.32	18.00	
	Median	10.00	90.50	5.32	18.00	
H. Medical Sciences	Max.	10.00	269.00	14.16	26.00	3
	Min.	6.00	82.00	4.08	19.00	
	Mean	8.00	152.33	7.44	21.67	
	Median	8.00	106.00	4.10	20.00	

Table 4. The distribution of the annual values of indicators

	Web of Science Documents	*Times Cited*	*Citation Impact*	*Cumulative Citations per Year*	*H-index*	*CNCI*	*Network Density*	*Clustering Coefficient*
Max.	183	18266	2512	9338	125	36.525	1	1
Min.	0	0	0	0	0	0	0	0
Mean	11.683	502.535	40.078	469.621	19.053	1.401	0.119	0.161
Median	6	166	22.824	140	13	0.978	0.083	0

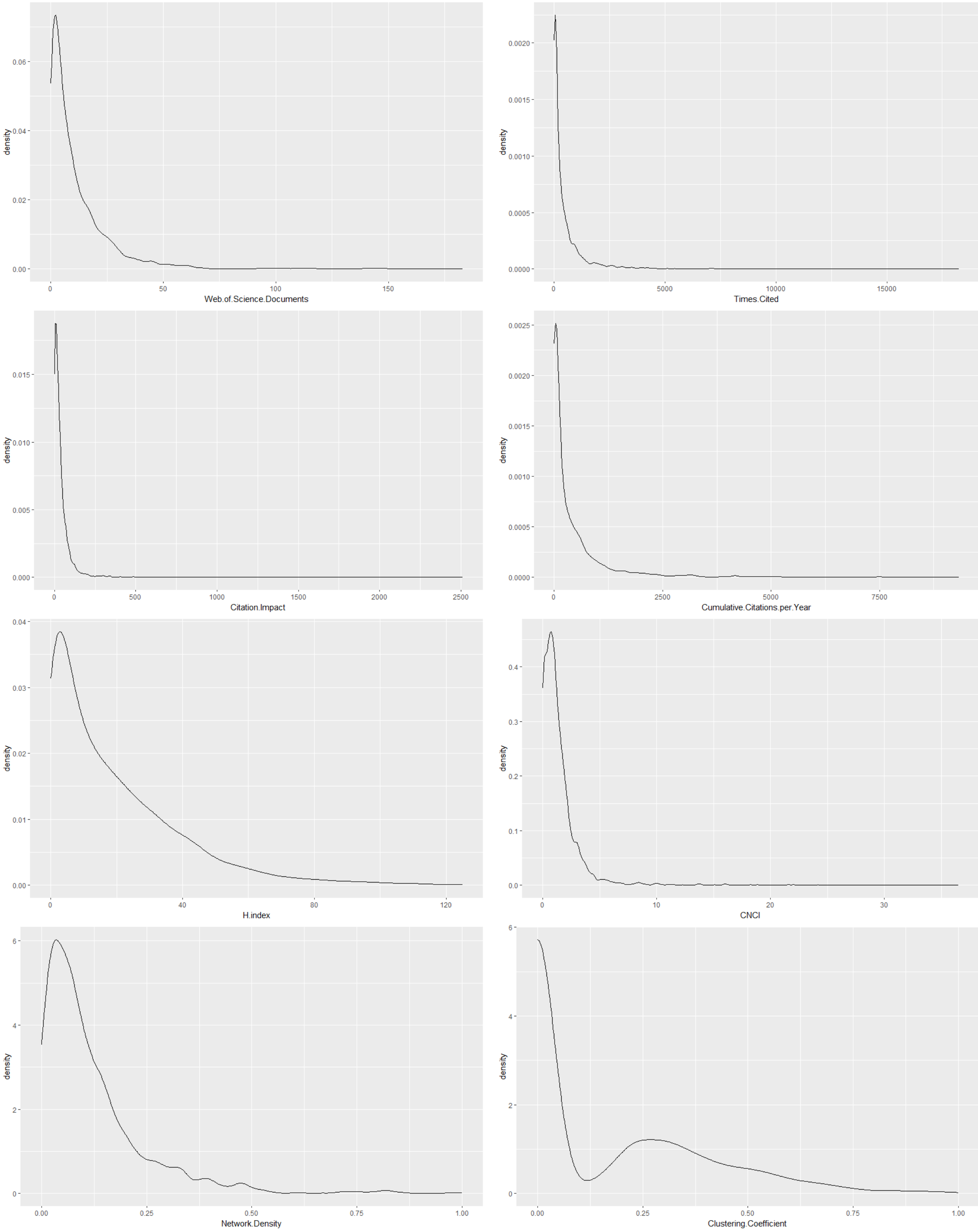

Figure 4. Density plot of the annual values of indicators for research teams

Calculate the annual team effectiveness for each team

Generally, principal component analysis (PCA) can be used to analyze cross-sectional data from multiple samples. However, our data also contains a temporal dimension, with each team having multiple years of indicator data. Thus, we need to apply the generalized principal component analysis (GPCA), which considers temporal dynamics based on the principles of PCA and uses panel data for dynamic analysis to reflect the dynamic characteristics and trajectories of all objects. In this study, we use SPSS Statistics to run the PCA for our samples. And we will normalize the data of indicators by Min-Max Normalization before this software performs the analysis. In the following analysis, we can first get the test results about the correlation matrix, the Kaiser–Meyer–Olkin (KMO) test, and Bartlett's test of sphericity, which are shown in Table 5 and Table 6. In Table 6, the KMO value is 0.681> 0.5, and Bartlett's test of sphericity is significant with a p-value of < 0.05. Thus, these results indicate that there is a strong correlation between the indicators, and the principal component analysis and factor analysis can be appropriate for the data.

Table 5. Correlation matrix between indicators

	Web of Science Documents	*Times Cited*	*Citation Impact*	*Cumulative Citations Per Year*	*H-Index*	*CNCI*	*Network Density*	*Clustering Coefficient*
Web of Science Documents	1.000	.590	.026	.565	.410	.080	.575	.529
Times Cited	.590	1.000	.421	.453	.296	.443	.431	.408
Citation Impact	.026	.421	1.000	.071	.036	.827	.090	.092
Cumulative Citations Per Year	.565	.453	.071	1.000	.828	.210	.304	.256
H-Index	.410	.296	.036	.828	1.000	.199	.234	.175
CNCI	.080	.443	.827	.210	.199	1.000	.100	.099
Network Density	.575	.431	.090	.304	.234	.100	1.000	.852
Clustering Coefficient	.529	.408	.092	.256	.175	.099	.852	1.000

Table 6. KMO and Bartlett's test of sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		*0.681*
Bartlett's Test of Sphericity	Approx. Chi-Square	14478.794
	Df	28
	Sig.	0.000

Furtherly, we can get the eigenvalues, cumulative variance contribution rates, and component matrix of the principal components, and they are shown in Table 7 and Table 8, respectively. We can know that three principal components are extracted. The cumulative variance contribution of these three principal components reached 83.449%, indicating that they contain 83.449% (>80%) of the information of the original 8 indicators, retaining adequate original information to be acceptable. Based on the data in Table 7 and Table 8, we can calculate indicator weights using principal component analysis. The step is as follows.

The linear combination coefficients matrix

The linear combination coefficients in principal component analysis (PCA) are the coefficients between the variables and the principal components, which are used to calculate the score of each principal component. Specifically, each principal component can be represented as a linear combination of indicators:

$$F_j = u_{1j}*ZX_1 + u_{2j}*ZX_2 + ... + u_{ij}*ZX_i \quad (1)$$

where F_j represents the j principal components, u_{1j} to u_{ij} represent the corresponding linear combination coefficients of each principal component, and ZX_1 to ZX_i represents the i standardized variables.

In this study, the component matrix shows the correlation relationship between the eight variables and the three principal components extracted. And the principal components are linear combinations of variables. Based on this formula,

$$\rho(ZX_i, F_j) = u_{ij}\sqrt{\lambda_j} \quad (2)$$

where ρ represents the coefficients (loadings) corresponding to each variable in the component matrix; F_j represents the principal components; ZX_i represents the standardized variables; u_{ij} represents the linear combination coefficients; and λ_j represents the eigenvalue of the jth principal component.

Thus, the linear combination coefficients can be calculated using the component matrix and eigenvectors of principal components in PCA. Specifically, the coefficients are equal to the loadings in the component matrix divided by the square root of eigenvalues of corresponding principal components. And then, the linear combination coefficients of each principal component can be computed. The results are shown in Table 9.

Therefore,

$$F_1 = 0.420*ZX_1 + 0.414*ZX_2 + 0.198*ZX_3 + 0.392*ZX_4 + 0.333*ZX_5 + 0.242*ZX_6 + 0.390*ZX_7 + 0.371*ZX_8 \quad (3)$$
$$F_2 = -0.233*ZX_1 + 0.169*ZX_2 + 0.644*ZX_3 - 0.113*ZX_4 - 0.105*ZX_5 + 0.616*ZX_6 - 0.230*ZX_7 - 0.218*ZX_8 \quad (4)$$
$$F_3 = -0.015*ZX_1 - 0.067*ZX_2 - 0.120*ZX_3 + 0.507*ZX_4 + 0.577*ZX_5 + 0.009*ZX_6 - 0.418*ZX_7 - 0.465*ZX_8 \quad (5)$$

Coefficients of variables in the composite score

Following the calculation of the scores for each of these principal components, weighting coefficients are assigned to each component based on its contribution to the variance, and the weights are added to obtain the composite score (F). And the formula is shown below.

$$F=a_1*F_1 + a_2*F_2 + \dots + a_i*F_i \quad (6)$$
$$F=w_1*ZX_1 + w_2*ZX_2 + \dots + w_n*ZX_n \quad (7)$$

Where a_i represents the ratio of the variance of the ith principal component to the cumulative variance of all principal components; w_n represents the coefficients of variables.

In this study, they can be expressed as follows.

$$F = (43.710/83.449)*F_1 + (22.635/83.449)*F_2 + (17.104/83.449)*F_3 \quad (8)$$
$$F=w_1*ZX_1+w_2*ZX_2+\dots+w_8*ZX_8 \quad (9)$$

When we take formulas (3), (4), and (5) into formula (8), we can calculate the coefficients of variables (w_1 to w_8) in formula (9), and the results are shown in Table 9.

Weights of the indicators in the Team Effectiveness Index

When we normalize the coefficients of variables in the composite score model, the weights of variables can be determined. And the results are shown in Table 9. Thus, we can give the formula about the team effectiveness index.

$$\text{Team Effectiveness Index (TEI)} = 0.097*X_1 + 0.157*X_2 + 0.160*X_3 + 0.175*X_4 + 0.166*X_5 + 0.186*X_6 + 0.035*X_7 + 0.025*X_8$$

Where X_1 to X_8 represent the standardized values of indicators.

Table 7. Eigenvalues and the variance contribution rate of the principal generalization component

Principal component	*Eigenvalues*	*Contribution rate/%*	*Cumulative contribution rate/%*
F_1	3.497	43.710	43.710
F_2	1.811	22.635	66.345
F_3	1.368	17.104	83.449

Table 8. Component matrix of generalization principal component

Indicators	*Loading*		
	1	2	3
Web of Science Documents (X_1)	.785	-.313	-.017
Times Cited (X_2)	.775	.227	-.078
Citation Impact (X_3)	.371	.866	-.140
Cumulative Citations Per Year (X_4)	.733	-.152	.593
H-Index (X_5)	.623	-.141	.675
CNCI (X_6)	.453	.829	.011
Network Density (X_7)	.730	-.309	-.489
Clustering Coefficient (X_8)	.693	-.293	-.544

Table 9. Component matrix of generalization principal component

Indicators	*Linear combination coefficients*			*Coefficients in composite score*	*Normalized weights*
	1	2	3		
Web of Science Documents (X_1)	0.420	-0.233	-0.015	0.154	0.097
Times Cited (X_2)	0.414	0.169	-0.067	0.249	0.157
Citation Impact (X_3)	0.198	0.644	-0.120	0.254	0.160
Cumulative Citations Per Year (X_4)	0.392	-0.113	0.507	0.279	0.175

H-Index (X_5)	0.333	-0.105	0.577	0.264	0.166
CNCI (X_6)	0.242	0.616	0.009	0.296	0.186
Network Density (X_7)	0.390	-0.230	-0.418	0.056	0.035
Clustering Coefficient (X_8)	0.371	-0.218	-0.465	0.040	0.025

Since the measure of this index consists of 3 dimensions, where X1 belongs to the team productivity dimension, X7 and X8 to the team collaboration dimension and the other 5 indicators to the team impact dimension, we can assume that these 3 dimensions together influence the change in team effectiveness. If we use the 3 indexes of team productivity, team impact and team collaboration to demonstrate team effectiveness, then we can have the following formulas after the weights have been normalized.

Team Productivity Index (TPI) = X_1
Team Impact Index (TII) = $0.186*X_2 + 0.189*X_3 + 0.208*X_4 + 0.197*X_5 + 0.221*X_6$
Team Collaboration Index (TCI) = $0.587*X_7 + 0.413*X_8$
Where X_1 to X_8 represent the standardized values of indicators.

Based on these four formulas, we can calculate the annual Team Effectiveness Index, Team Productivity Index, Team Impact Index, and Team Collaboration Index for each research team.

Identification of the life cycle stages of research teams

Change points detection to the curves of the Team Effectiveness Index
It is challenging to quantify changes in team effectiveness since the values of the indicators normalized by the Min-Max Normalization have a range of values between 0 and 1. Thus, we want to ensure that the indicators have a range of values from 0 to 100. That is to say, we apply the following formula to transform our data about indicators.

$$X' = ((X_i - X_{min})/(X_{max} - X_{min}))*100$$

Where X_i represents the original values of indicators; X_{max} and X_{min} represent the maximum and minimum values of X_i; X' represents the transformed values of indicators.

In addition, the transformed values of the indicators are averaged over a 3-year moving average to lessen the effect of outliers on the curve's trend. After these steps to calculate, we can obtain the curves of the team effectiveness index for each research team. And we only give two cases (in Figure 5) in the following analysis due to the limited space.

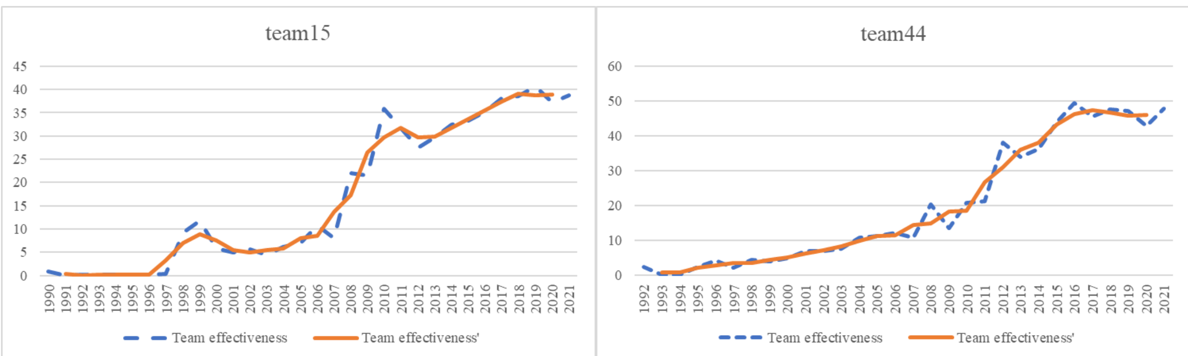

Figure 5. The annual team effectiveness index of research teams

Tips: the curves of team effectiveness use the normalized values of indicators, and the curves of team effectiveness use their 3-year moving average values to calculate.

And then, we use the package 'strucchange' in R to detect the change point in these curves. Based on the 3-year moving average of the team effectiveness index, this tool detects three change points for these two teams respectively. And 1996, 2007, 2014 are the years with significant changes for team15, and 2002, 2009, 2013 are the years for team44.

Identification of the life cycle stages of research team development

Furtherly, we can calculate the average and change rate of the team effectiveness index in different phases of team development. According to the rules for classifying and judging the life cycle stages of research teams, we can know that from the first to the fourth stage, the average team effectiveness ranges from small to large, and the growth rate ranges from small to large and then becomes smaller, with the largest growth rate in the second stage. For team 15, the change in the average and growth rate of its team effectiveness across the four stages is consistent with the basic pattern of team life cycle development. Therefore, it may be deemed to have progressed through the four stages in sequence.

For team 44, the average team effectiveness has an increasing average value. However, the first phase has the largest growth rate. Compared to the second phase, the first phase has a large growth rate and a lower average of team effectiveness. Thus, it may be the second or third life cycle stage. In addition, the second, third, and fourth phases have continued the trend of growth, and there has been no decline or cessation of growth. Therefore, the first phase can be considered to be the second life cycle stage. Accordingly, the second phase can be considered the third life cycle stage of team development. However, when the second, third, and fourth phases are analyzed together, they correspond to the basic pattern for the development of teams' life cycle stages. This team can therefore be considered to have undergone a double S-curve life cycle development. The relevant results are shown in Table 10.

Table 10. Identification of the life cycle stages of research team development

Team15				*Team44*			
Phases	*Average*	*Change rate*	*Life cycle stages*	*Phases*	*Average*	*Change rate*	*Life cycle stages*
1990-1996	0.178	0.079	First	1992-2002	3.730	0.304	Second
1996-2007	7.152	1.432	Second	2002-2009	12.675	0.147	Third (First)
2007-2014	28.044	0.143	Third	2009-2013	28.158	0.195	Second
2014-2021	37.184	0.034	Fourth	2013-2021	44.868	0.036	Third

Comparative analysis

To distinguish the method proposed in this paper and the previous method used to identify team life cycle stages, we analyzed the curves of team development with different approaches. Since there are only a few references about the use of quantitative bibliometric indicators to identify the life cycle stages of teams, and they identify the different stages of team development only by the number of team publications, we will compare the differences in the results derived from these two approaches. We use the R software to detect the change points in the curves of the Team Productivity Index and Team Effectiveness Index for two case teams, respectively. The change points detected are shown in Table 12.

Table 11 shows that these two approaches produce nearly identical results for team 44. However, there is a significant difference between the two findings for team 15, despite the fact that they all identify three change points. That is, the trends in these two types of curves over time vary and are not identical. As a system, research teams' activities are dynamic and involve a number of dimensions. If only one dimension of team development is observed to analyze the change of team life cycle stages, there will undoubtedly be some limitations. According to our findings, the life cycle stages of research teams identified by the team publications and team effectiveness index are sometimes very similar, but there are also teams whose productivity and overall team effectiveness trends are very different. If the latter, it should be judged by the team effectiveness index, since this index combines the changes in team development in multiple dimensions. For example, sometimes the number of publications produced by team members working together stays at a plateau, but the quality of the papers produced continues to improve. In this instance, the team's impact is increasing while its productivity stays stable, and the team's effectiveness is actually increasing.

Table 11. Change points for curves of the TPI and TEI

Team15			*Team44*		
Team effectiveness	*Team productivity*	*Life cycle stages*	*Team effectiveness*	*Team productivity*	*Life cycle stages*
1990-1996	1990-1999	First	1992-2002	1992-2001	Second
1996-2007	1999-2003	Second	2002-2009	2001-2009	Third (First)
2007-2014	2003-2007	Third	2009-2013	2009-2013	Second
2014-2021	2007-2021	Fourth	2013-2021	2013-2021	Third

Conclusion and discussion

In this study, we propose a new method to identify and understand the life cycle development of research teams. Firstly, we have refined Tuckman's team development model and built a new conceptual life cycle model for research team development. This model depicts the curve of team effectiveness over time and divides the various stages of the team life cycle according to the inflection points of the curve. Based on this conceptual model, we can gain a macro understanding of the development of research teams' life cycle stages. Then, we select several bibliometric indicators to measure team effectiveness from three dimensions, including team productivity, team impact, and team collaboration, and apply the generalized principal component analysis (GPCA) to develop the Team Effectiveness Index (TEI). Accordingly, we also get the Team Productivity Index (TPI), Team Impact Index (TII), and Team Collaboration Index (TCI). After that, in order to get the curves of team effectiveness over time, we calculate the annual values of the index. And then, we use the package 'strucchange' in R to detect the change points in the curves of TEI, and these points can divide the life cycle of research teams into multiple phases. We finally calculate TEI's average and change rate at different phases and determine which team life cycle stages these phases detected should belong to. Based on the identification results about the life cycle stages of research teams, we can further understand research team development. We analyzed research teams' development through the case study and tested our method. Additionally, we compared the findings acquired by our method to those obtained by the previous method. The data reveal that the two methods do not always give the same identification results for teams' life cycle stages. In comparison to the previous method, our method has broader applicability because it incorporates more aspects of team development.

From the analysis results, we can know our method is applicable to both single S-curve growth patterns and multiple S-curve growth patterns of team life cycle development. And the multiple

S-curve models can be used to identify the fluctuating life cycle development. Although our method is promising, there are also some limitations. Firstly, we consider those project participators as one virtual research team and their collaborative publications as team outcomes in order to observe the full and true life cycle of teams. This has a certain legitimacy since team members contain the most core members. However, it may result in some members not being identified, as there may be a few core collaborators who do not appear in the list of project members. And it may lead to a slight reduction in the number of team publications. Secondly, our method only can be used to research teams in the field of basic research. Because they usually publish their research findings in papers, we can use bibliometric indicators and data to measure the team's effectiveness or performance. Generally, team effectiveness can involve some other dimensions beyond publications. For example, teams may produce some patents. We also should calculate the number of students that teams have trained and how many awards the teams received. All these should ideally be considered in the calculation of team effectiveness. Thirdly, the Team Effectiveness Index is ‘lagging’, since the accumulation of several bibliometric indicators (such as citations.) takes time to increase. These indicators cannot show the changes in team effectiveness in real-time. So, we should try some new indicators to measure team effectiveness in future studies. Finally, while Principal Component Analysis (PCA) can reduce high-dimensional data to low-dimensional data and help users remove the redundant information in the data, it is too data-dependent and sensitive to outliers since outliers can greatly affect the principal components. This will have some influence on the construction of the Team Effectiveness Index.

Overall, our method only provides one possibility for identifying and understanding the team life cycle development. There are also many questions that should be analyzed and discussed in detail in future studies. Firstly, the stages of team development in the conceptual models, both in our study and in the previous studies, are ideal representations of teams’ situations and team’s “complete” developmental life cycle, and actually, teams do not necessarily have all lifecycle stages and all the characteristics of each stage in practice (Edison, 2008; Jin, Qian, & Shao, 2006). Secondly, there may be no clear boundaries or crossings between two adjacent life cycle stages, and they may even have the characteristics of both life stages when teams enter a stage transition (Jin et al., 2006).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (Grant Nos. 71974150, 72004169), and the National Laboratory Center for Library and Information Science at Wuhan University.

References

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. PLoS One, 4(6), e6022.
Braam, R., & Van den Besselaar, P. (2010). Life cycles of research groups: the case of CWTS. Research evaluation, 19(3), 173-184.
Braam, R., & van den Besselaar, P. (2014). Indicators for the dynamics of research organizations: a biomedical case study. Scientometrics, 99(3), 949-971.
Clarivate. (2021). Indicators Handbook. Retrieved from https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-about.htm
Collins, C. G., & Parker, S. K. (2010). Team capability beliefs over time: Distinguishing between team potency, team outcome efficacy, and team process efficacy. Journal of Occupational and Organizational Psychology Review, 83(4), 1003-1023.

Curry, S., de Rijcke, S., Hatch, A., Pillay, D., van der Weijden, I., & Wilsdon, J. (2022). The changing role of funders in responsible research assessment: progress, obstacles and the way ahead Research on Research Institute. Report. In.

Edison, T. (2008). The team development life cycle. Defense AT&L, 37(3), 14-17.

Franceschet, M. (2009). A cluster analysis of scholar and journal bibliometric indicators. Journal of the American Society for Information Science
and Technology, 60(10), 1950-1964.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. Science, 122(3159), 108-111.

Gersick, C. J. (1988). Time and transition in work teams: Toward a new model of group development. Academy of Management journal, 31(1), 9-41.

Giordani, P. (2018). Principal Component Analysis. In R. Alhajj & J. Rokne (Eds.), Encyclopedia of Social Network Analysis and Mining (pp. 1831-1844). New York, NY: Springer New York.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. Annual Review of Psychology, 47(1), 307-338.

Hurt, A. C., & Trombley, S. M. (2007). The Punctuated-Tuckman: Towards a New Group Development Model. Online Submission.

Jin, H., Qian, Y., & Shao, J. (2006). Team life cycle theory and relevant research progress. Science & Technology Progress and Policy(07), 194-196.

Kulikowski, K., Przytuła, S., & Sułkowski, Ł. (2022). When publication metrics become the fetish: The research evaluation systems' relationship with academic work engagement and burnout. Research evaluation. Retrieved from https://doi.org/10.1093/reseval/rvac032

Nane, T., & Bachasingh, A. (2019). Dependence modeling of bibliometric indicators with copulas. Paper presented at the 17th International Conference on Scientometrics and Informetrics, ISSI 2019.

National Research Council. (2015). Enhancing the effectiveness of team science. Washington (DC): National Academies Press.

Peralta, C. F., Lourenço, P. R., Lopes, P. N., Baptista, C., & Pais, L. (2018). Team development: definition, measurement and relationships with team effectiveness. Human Performance, 31(2), 97-124.

Potter, I. (2016). Beyond the impact factor: taking a wider view of journal evaluation. Insights, 29(1).

Price, D. J. d. S. (1963). Little science, big science: Columbia University Press.

Rey-Rocha, J., Garzon-Garcia, B., & Martín-Sempere, M. J. (2006). Scientists' performance and consolidation of research teams in Biology and Biomedicine at the Spanish Council for Scientific Research. Scientometrics, 69(2), 183-212.

Shuffler, M. L., Salas, E., & Rosen, M. A. (2020). The evolution and maturation of teams in organizations: Convergent trends in the new dynamic science of teams. Frontiers in Psychology, 11, 2128. Retrieved from https://doi.org/10.3389/fpsyg.2020.02128

Towler, A. (2020). Team effectiveness and team performance: What it is and how to improve it. Retrieved from https://www.ckju.net/en/dossier/team-effectiveness-and-team-performance-what-it-and-how-improve-it

Tuckman, B. W. (1965). Developmental sequence in small groups. Psychological bulletin, 63(6), 384-399.

Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. Group & organization studies, 2(4), 419-427.

Wikipedia. (2023). Principal component analysis. Retrieved from https://en.wikipedia.org/wiki/Principal_component_analysis

Wu, B. (2019). WHU's Center of the Studies of Information Resources listed as Innovative Research Group Project of NSFC. Retrieved from https://en.whu.edu.cn/info/1081/4611.htm