J Vis
https://doi.org/10.1007/s12650-022-00862-7

REGULAR PAPER

Teng Wang · Zeyu Li · Jiawan Zhang

Egocentric visual analysis of dynamic citation network

Received: 27 March 2022 / Revised: 23 May 2022 / Accepted: 1 June 2022

Abstract Egocentric citation network visualization focuses on the citation relationship around a specific node, which can help gain insights into citation patterns clearly. Previous techniques can effectively analyze the whole citation network at a single level, but cannot gain the temporal citation patterns of a specific field at different levels and especially analyze the authors that have promoting influence on the citation relationships, which causes the incomprehensive perspective on understanding the citation relationships. In this paper, we design and implement a novel egocentric visual analysis system of citation network. In order to perform multiple-level analysis for the citation relationships between a specific field and other fields, we construct a new hierarchical egocentric network. Based on the vivid pollen-spread metaphors which can interpret the constructed network impressively, the citation relationship is similar to the "pollen spread" between "flowers," and each author that has promoting influence on the citation relationships plays a role of "bee." We provide abundant visualizations and interactions by these metaphors, which can effectively obtain the temporal patterns of the citation network. Finally, we evaluate the effectiveness of our system through case studies and user study.

Keywords Citation network · Dynamic network · Egocentric network · Visual analysis

1 Introduction

As more and more academic data can be available, citation network analysis has become a popular research topic that can reveal meaningful citation patterns. Citation network is a typical dynamic network, which contains noteworthy temporal information. Among a lot of analytical techniques of dynamic network, visualization offers an intuitive perspective to help gain hidden patterns (Beck et al 2017). Compared with visualizing the whole network, egocentric network visualization focuses on the local subnetwork that represents the relationship between a focal individual and others connected to it, which can display the relational pattern of the focal node clearly. The effectiveness of egocentric dynamic network visualization has been demonstrated (Bryan et al 2013; Liu et al 2015; Shi et al 2015; He et al 2016; Cao et al 2016; Wu et al 2016; Zhao et al 2016; Law et al 2018; Chen et al 2019b).

T. Wang · Z. Li · J. Zhang (✉)
College of Intelligence and Computing, Tianjin University, Tianjin, China
E-mail: jwzhang@tju.edu.cn

T. Wang
E-mail: wt0201@tju.edu.cn

Z. Li
E-mail: lzytianda@tju.edu.cn

Published online: 20 July 2022

Many egocentric visualization techniques of dynamic network can be used to analyze citation network. From the perspective of analysis level, these techniques can be divided into single-level and multiple-level. Single-level techniques usually visualize a single network at a time (e.g., (Liu et al 2015; Shi et al 2015; Zhao et al 2016)). Because of lacking hierarchical structure, these techniques cannot offer progressive exploration which is beneficial to the cognitive process of analysts, and many valuable findings at different analysis levels are ignored. There are also a few multiple-level visual analysis techniques of egocentric dynamic network (e.g., (He et al 2016; Wu et al 2016)). They usually choose to visualize all egocentric networks at the highest level and a single egocentric network at the lower level. For obtaining the general temporal patterns clearly, it is necessary to analyze the egocentric dynamic network where each node represents a set of entities, such as the citation network between a specific field and other fields. However, previous multiple-level techniques cannot effectively analyze this kind of network, because they focus on the egocentric network where each node represents a single entity. Besides, preview studies of citation network visualization also have limitations in terms of concrete tasks. They do not place emphasis on the authors who have promoting influence on the citation relationships, which misses a new perspective on understanding the citation relationships.

To address the above concerns, we design and implement a novel egocentric visual analysis system of citation network. In order to perform multiple-level analysis for the citation relationships between a specific field and other fields, we construct a new hierarchical egocentric network. Based on the vivid pollen-spread metaphors which can interpret the constructed network impressively, the citation relationship is similar to the "pollen spread" between "flowers," and each author that has promoting influence on the citation relationships plays a role of "bee" that contributes to the "pollen spread." In order to obtain the temporal patterns of the citation network, we provide abundant visualizations and interactions, such as the elaborate visual encoding for the citation network by metaphors. To evaluate the effectiveness of our system, we conduct case studies and user study. In the case studies, we use our system to analyze the citation network between the visualization field and other fields in Computer Science, which can get a few meaningful discoveries.

Our main contributions in this paper are summarized as follows: (1) We design and implement a novel egocentric visual analysis system of citation network based on the pollen-spread metaphors, which can perform multiple-level analysis for the dynamic citation relationships between a specific field and other fields. (2) We get a few discoveries that are worthy of concern by using our system to analyze the citation network between the visualization field and other fields in Computer Science.

2 Related works

2.1 Egocentric network visualization

Egocentric visualization can effectively analyze citation network. From the perspective of the analysis level, the techniques of egocentric network visualization can be divided into single-level and multiple-level.

Many studies focus on egocentric network at a single level and visualize a single network at a time. The 1.5D visualization design (Shi et al 2015) encoded the time information of each edge by the location of its endpoint, which can reduce the visual complexity of the egocentric dynamic network. Compared to the 1.5D approach, EgoNetCloud (Liu et al 2015) introduced simplification methods and adopted an improved stress majorization approach, which can effectively analyze event-based egocentric network. EgoNetCloud used a cloud metaphor for visualization design. Besides, a subway-based metaphor (Zhao et al 2016) has been used to visualize egocentric network. These studies can effectively achieve given analysis goals. However, due to their limited analysis ability, it is difficult for them to explore egocentric networks at multiple levels. Compared with these studies, our system can analyze egocentric citation network at different levels.

For the exploration of egocentric network at different levels, several multiple-level visual analysis approaches have been proposed. D-Maps+ (Chen et al 2019a) supported egocentric analysis of user behaviors and diffusion patterns in social media, which can effectively explore single user and group users in information diffusion. In D-Maps+, a novel design based on the map metaphor was introduced to visualize the aggregated information diffusion process intuitively. For the egocentric network of player in game, MMOSeer (Li et al 2017) provided a holistic exploration, including the overview of the entire intimacy network, the evolution of the egocentric network, and the detailed activities in game. Additionally, some approaches have been proposed with similar level design. DMNEVis (Peng et al 2018) treated the egocentric network as a part of the group network and explored the evolution of the dynamic multivariate network from group network to egocentric network step by step. egoSlider (Wu et al 2016) revealed the evolutionary patterns of egocentric networks at three levels: the macroscopic level for summarizing all egocentric networks, the mesoscopic level for overviewing specific individuals' egocentric network evolutions, and the microscopic level for displaying detailed temporal information in a single egocentric network. In a way, MENA (He et al 2016) constructed the analysis levels that were similar to egoSlider. These three techniques can comprehensively explore egocentric networks, but they cannot effectively analyze the general patterns from an egocentric perspective. And they did not especially explore the noteworthy details on the edges. In our system, we focus on multiple-level analysis and can gain the general citation patterns from an egocentric perspective. Moreover, we design abundant visualizations for the edges in citation network.

2.2 Dynamic network visualization

Citation network is a typical dynamic network. Many visualization techniques have been proposed to analyze dynamic network, which can be divided into animation-based, timeline-based, and hybrid.

Animation is an effective way to show the changes of a dynamic network. Galex (Li et al 2020) used animation to show the dynamics of hotspots in different academic networks, which can effectively reveal research trends. Three strategies were explored to stage animations for online dynamic networks, including time-based, event-based, and a new hybrid approach (Crnovrsanin et al 2021). The animation-based approaches mainly focused on the time-varying topology structure of a network. Because they cannot show networks at all time steps simultaneously, analysts need to memorize the status of the network and probably suffer from cognitive load. As a result, we do not use animation to analyze the evolution of citation network in our system.

Compared with animation-based approaches, timeline-based approaches visualize the information in dynamic network with chronological alignment. Whisper (Cao et al 2012) used a sunflower metaphor to trace the spatiotemporal process of information diffusion in social media and revealed the specific relationships among the users by timeline. GraphFlow (Cui et al 2014) proposed a static flow visualization to summarize the graph metrics of the entire graph and its evolution over time. Flow visualization can also be used to visualize the evolution of communities in dynamic networks (Vehlow et al 2015). Besides, several visualization designs have been proposed, such as TimeRadarTrees (Burch and Diehl 2008), TimeArcTrees (Greilich et al 2009), and TimeArcs (Nhon et al 2016). Most of timeline-based approaches explored dynamic network at a single level. In our system, we enable analysts to obtain temporal information at multiple levels and visualize the details of each citation relationship by timeline.

Several approaches have combined animation and timeline, such as DiffAni (Rufiange and McGuffin 2013) which visualized the network as a sequence of diff tiles, animation tiles, and small multiple tiles. Hybrid approaches sometimes have advantages over non-hybrid techniques. However, animation and timeline are not entirely complementary. The limitations of animation and timeline mentioned above cannot be effectively overcome by hybrid approaches.

2.3 Citation network visualization

From the perspective of node granularity, the techniques of citation network visualization can be divided into visualizations with or without paper aggregation.

A few visualizations have been proposed to analyze the citation relationships without paper aggregation. For the papers presented at InfoVis, the citing motivations and topics of citations have been explored (Yoon et al 2020). Besides, GraphAEL (Erten et al 2003) has provided an animated visualization for dynamic citation network, which enables the identification of changes while preserving the mental map of analysts. Because these visualizations focus on the detailed citations, many findings about the general citation patterns are ignored. In our system, we do not directly visualize the citation network where each node represents a single paper.

There are also many visualizations which focus on the citation relationships with paper aggregation. In CiteRivers (Heimerl et al 2016), the citations aggregated by user steer are linked to the content of the citing publications using interactive visualization. What's more, the citation patterns of the physics papers aggregated by journals and journal volumes have been shown by citation flows (Herr et al 2008). These visualizations can display the general information of the citation network. However, most of them cannot perform multiple-level visual analysis for the egocentric citation network where each node represents the paper set of a field. Although Influence Map (Shin et al 2019) can show the aggregated influence around a given academic entity by a new visual metaphor, the time-varying citation number in each citation relationship cannot be explored. What's more, they did not especially analyze the authors who have promoting influence on the citation relationships. In our system, these meaningful explorations can be achieved.

3 Requirement analysis

In order to obtain clear requirements for helping analysts explore the citation patterns of a specific field, we interviewed three experts who are researchers in the field of visualization. The identified requirements are described as follows.

R1: Gaining an overview of the citation relationships between a specific field and other fields. Analysts want to perform the egocentric analysis so that they can clearly focus on the citation relationships about a specific field. To gain a basic understanding about the citation patterns of the specific field, analysts want to find all related fields that have the citation relationships with the specific field as well as the main subfields of each field.

R2: Exploring the attributes of the fields/subfields in the citation relationships. Analysts want to explore several attributes of each field, including the paper number, the citing number, and the cited number. They want to analyze the closeness of the citation relationships based on the paper number, and compare the citing number and the cited number of each field. For each subfield, analysts are interested in the time span of the papers.

R3: Analyzing the strength evolution in the citation relationships. Analysts want to learn how the citation strength changes over time and analyze the trends that are worth considering. Besides, analysts want to compare the strength evolution in different citation relationships.

R4: Finding the influential authors in the citation relationships of different years. Analysts want to learn how the authors influence the citation relationships so that they can have a new perspective on understanding the citation patterns. For the influential authors in different years, analysts want to explore their temporal patterns.

R5: Mining the meaningful patterns of multiple-level citation relationships. Because the citation relationships about a field are formed by aggregating the citation relationships about its subfields, hierarchical analysis is necessary. Analysts want to analyze different kinds of citation relationships progressively, such as: the citation relationship between the specific field and all related fields, the citation relationship between the specific field and each related field. These citation relationships contain the citation patterns at different levels. Besides, analysts want to analyze the citation relationships influenced by a specific author so that they can learn more about this author.

These requirements are related to multiple aspects of citation relationships. R1, R2, and R3 have been involved in previous studies. R4 and R5 are unique requirements. For R5, multiple-level citation relationships will be analyzed, which enables analysts to obtain comprehensive information when meeting R1, R2, and R3.

4 Visual analysis system

4.1 Overview

The overview of our visual analysis system is shown in Fig. 1. To gain the citation patterns of a specific field at different levels, we construct a new hierarchical egocentric network. The constructed network has two levels. We define the high level as the first level and the low level as the second level. Based on the pollen-spread metaphors, each paper set of a field or a subfield is represented by a "flower terrace" or a "flower," and each citation relationship is represented by a "pollen spread pipeline." In addition, each author that has promoting influence on the citation relationships is represented by a "bee." In the interface of our system, the focal "flower terrace" is on the left and the related "flower terraces" are on the right, while the main "flowers" in each "flower terrace" are also shown (R1). The "scutcheons" in each "flower terrace" show different attributes of the field, and the count of the "petals" in each "flower" represents the time span of the papers in each subfield (R2). Each "pollen spread pipeline" is in the middle of the interface (R3). For the

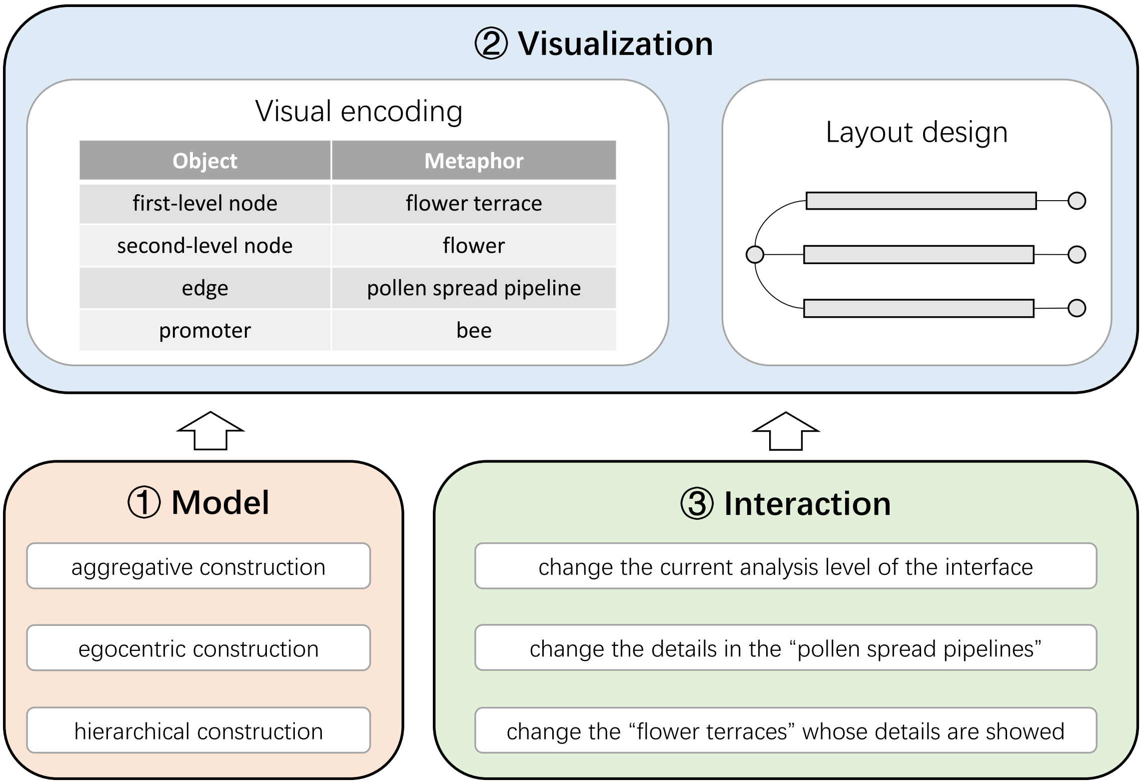


Fig. 1 The overview of our visual analysis system

"bees" representing authors, the interface shows the most influential "bees" of different time spans in each "pollen spread pipeline" (R4). And we define four analysis levels of the interface to analyze different citation relationships, which can be changed by interactions (R5). In addition, a few other interactions are designed for analysts.

4.2 Model

A citation network without paper aggregation can be modeled by $G = (V,E)$, where each node in the node set V represents a single paper and each edge in the edge set E represents the citation relationship from a paper to another paper. Each paper has the corresponding authors that have promoting influence on the citation relationships. These authors can form the set of promoters P in the edges. Because the citation network is directed, we expressly classify two types of promoters: the active promoters P_{active} and the passive promoters $P_{passive}$. For the citation relationship from a paper to another paper, the authors of the citing paper are defined as P_{active}, while the authors of the cited paper are defined as $P_{passive}$.

From this citation network, it is difficult to gain the citation patterns of a specific field at different levels. As a result, we need to construct a new hierarchical egocentric citation network. An example is shown in Fig. 2.

Aggregative construction. By the corresponding fields of all papers, we divide the node set V into n sets V_i' $(i \leq n)$. We treat each node set V_i' as a new node at the first level. On the basis of the previous network, we get the new edges and the promoters in these new edges. After treating the weight of each previous edge as 1, the weight of the new edge from a new node v_1 to another new node v_2 equals the weight sum of the previous edges from the previous nodes in v_1 to the previous nodes in v_2. Each promoter in the new network probably appears multiple times in a new edge. We define the times as the influence of the promoter. For an edge from a new node v_1 to another new node v_2, the influence of an active promoter can be computed by the times that he is an active promoter in the previous edges where the source node is in v_1 and the target node is in v_2, the influence of a passive promoter can be computed by the times that he is a passive promoter in the previous edges where the source node is in v_1 and the target node is in v_2. Because the previous network is dynamic, the weight of each new edge and the influence of each promoter are both time-varying.

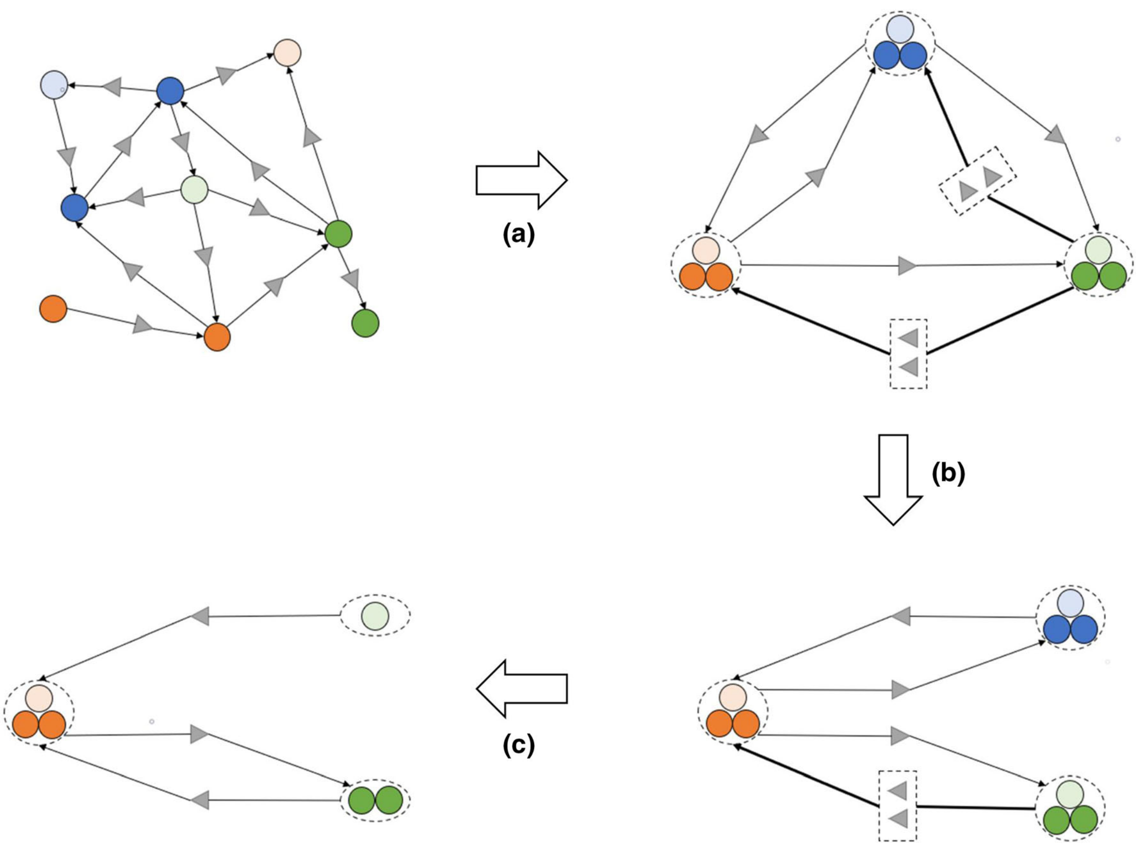


Fig. 2 An example that shows how a new citation network is constructed. Each colored circle is a node. A directed edge connects two nodes, and the promoters are shown as triangles. All nodes are aggregated to form multiple sets which are treated as new nodes, while the new edges and the promoters in these new edges are gained based on the previous network (**a**). For the egocentric analysis, the focal node and the related nodes are established (**b**). The related nodes at another level can be observed (**c**)

Egocentric construction. For egocentric analysis, we need to establish the focal node and the related nodes from the citation network. The focal node is the node that analysts want to explore, and the related nodes are the nodes that are connected to the focal node. We only focus on the edges between the focal node and the related nodes.

Hierarchical construction. For gaining the nodes at another level, each node set V'_i which represents a first-level new node can be divided into *m* sets V''_{i-j} ($j \leq m$) by the corresponding subfields of all papers. We treat each node set V''_{i-j} as a new node at the second level. The new edges and the promoters in these new edges can be gained by the same way.

4.3 Visualization

4.3.1 Visual encoding

In the constructed network, each paper set of a field/subfield is treated as a node, each citation relationship is treated as an edge, and each author is treated as a promoter. In order to analyze this network, we design several visualizations.The overall design is based on the pollen-spread metaphors which are shown in Fig. 3. In the natural world, pollen spreads from a flower to another flower, which is promoted by bees. In order to describe the process of pollen spread clearly, we assume that there are pollen spread pipelines between different flowers. Besides, the flowers of the same type can form a flower terrace, which can reflect hierarchical structure. To some extent, the pollen spread is similar to the citation relationship. In the citation network, each node can be treated as a "flower" or a "flower terrace," while each edge can be treated as a "pollen spread pipeline." For the promoters, we imagine them as "bees."

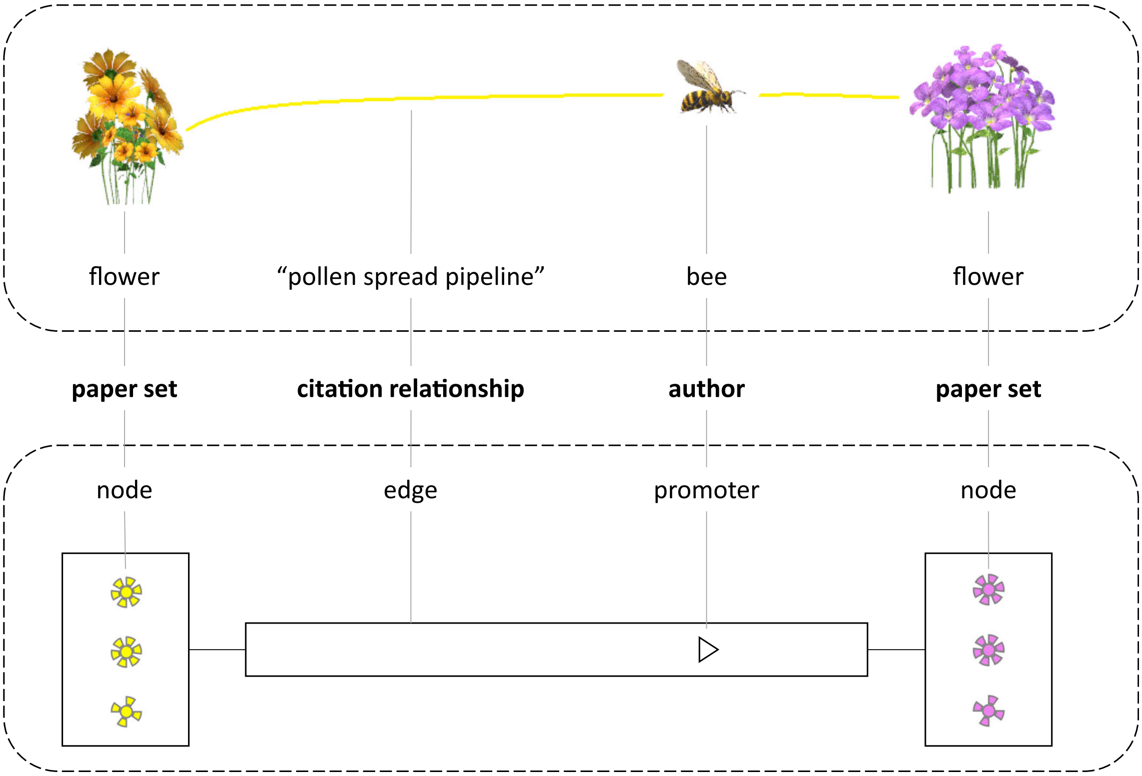


Fig. 3 The illustration of visual metaphors. The process that a flower spreads pollen to another flower is similar to the process that the papers in a set cite the papers in another set. The bees that promote the pollen spread are similar to the authors

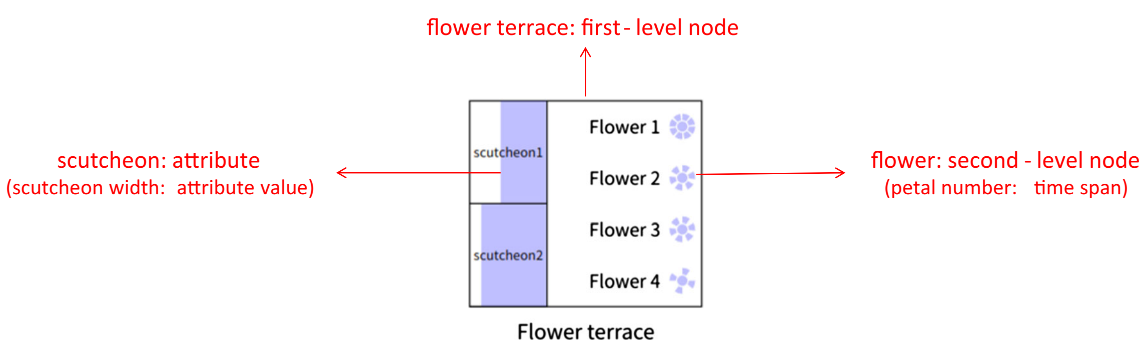


Fig. 4 The visual encoding for node. Each first-level node is treated as a "flower terrace," while each second-level node is treated as a "flower." The "scutcheons" in a "flower terrace" show different attributes of the first-level node

Node (flower terrace/flower). For the citation networks that have two-level hierarchical structure, each node at higher level is treated as a "flower terrace" and each node at lower level is treated as a "flower." The visual encoding for node is shown in Fig. 4. Each "flower terrace" is visualized by a rectangle. Because it is not practical to visualize many "flowers" in each "flower terrace," we decide to visualize the representative "flowers" which have high frequencies. In addition, we design the "scutcheons" which show different attributes of the "flower terrace." The breadth of the colored rectangle in each "scutcheon" represents the relative value for an attribute of the "flower terrace." For the focal "flower terrace" and the related "flower terraces" in the egocentric network, we use the same visual encoding. Each "flower" consists of a "stamen" visualized by a circle and several "petals" visualized by arcs. The count of the "petals" represents the time span of the "flower." The "flower" will have more "petals" if the time span is longer, which is similar to the natural law. The maximal count of the "petals" P_{max} and the maximal time span T_{max} can be set by analysts. The count of the "petals" P in a time span T can be gained as follows:

$$P = \begin{cases} P_{max}, & T \geq T_{max} \\ \left\lceil \frac{T}{T_{max}} * P_{max} \right\rceil, & T < T_{max} \end{cases}$$

Edge (pollen spread pipeline). Each edge is treated as a "pollen spread pipeline" which is visualized by a rectangle. For the edges in the constructed dynamic network, we emphasize the evolution of details, such as edge weight. We leverage stream chart to visualize the evolution of edge weight, which can be treated as "pollen stream." In the chart, the *x*-axis gives expression to the time, while the *y*-axis gives expression to the weight. The same scale is used for each stream chart so that analysts can compare the weights of different edges. Compared with other charts including bar chart and line chart, stream chart is in keeping with the metaphor of "pollen stream" better and can display the tendency smoothly (Fig. 5). If several "pollen streams" from different sources need to be visualized in a "pollen spread pipeline," we will use stacked stream chart (Byron and Wattenberg 2008). What's more, it is worth noting that the citation direction is the opposite direction of "pollen spread." If p_1 cites p_2, the "pollen" is considered to spread from p_2 to p_1.

Promoter (bee). The promoters influence each edge, and they are treated as "bees" vividly. Because of limited screen space, complex visualization design is not appropriate for the "bees." The optional shapes include circle, rectangle, triangle, and so on. Among these basic shapes, we choose triangle because this shape can imply the direction of the influence and analysts can associate the triangle with the "thorn" of a "bee." For the active "bees" and the passive "bees," we use the same visual encoding. The text information of each "bee" is at the side of each triangle.

4.3.2 Layout design

For the nodes represented by "flower terraces" and the relevant edges represented by "pollen spread pipelines," we have considered two layouts: the paralleled layout and the circular layout (Fig. 6). The paralleled layout is similar to some visualizations for bipartite networks (Steinböck et al 2018; Sun et al 2019). In this layout, the focal "flower terrace" is on the left of the interface, while the related "flower terraces" are on the right of the interface. The "pollen spread pipelines" are paralleled. In the circular layout, the focal "flower terrace" is in the middle, while the related "flower terraces" are around the focal "flower terrace." Although the circular layout preferably conforms to the meaning of egocentric analysis and has been used in some visualizations (Bryan et al 2013; Law et al 2018), it is difficult for analysts to contrast the details in radial "pollen spread pipelines." Hence, we choose to use the paralleled layout.

For the "bees" representing promoters, we visualize them in each "pollen spread pipeline" (Fig. 7). The influence of each "bee" is time-varying. It is impossible to visualize all "bees" at the same time. As a result, we decide to visualize the most influential "bees" in different time spans. Firstly, we divide the entire time into n approximately average time spans $[T_1,...,T_n]$. Then, we get the set of "bees" $B(T_i)$ in each time span T_i $(i \leq n)$ and sort them by their influence. Each "bee" b has an influence rank r_b. We set the maximal rank m and gain the most influential "bees" in T_i by $\{b \mid \forall\, b \in B(T_i),\ r_b \leq m\}$. Inspired by Sankey diagrams (Jiang and Zhang 2016), we visualize the "bees" in different parts of each "pollen spread pipeline." The position of each "bee" b that is shown depends on the time span T_i and its influence rank r_b. For each "bee" that appears in two sequential time spans, we use smooth curve to connect.

For the whole interface of our system, we define four analysis levels with different layout details. In the first-level interface which focuses on the focal "flower terrace" and all related "flower terraces," all related "flower terraces" are visualized on the right of the interface. In the second-level interface which focuses on the focal "flower terrace" and each related "flower terrace," only several related "flower terraces" are

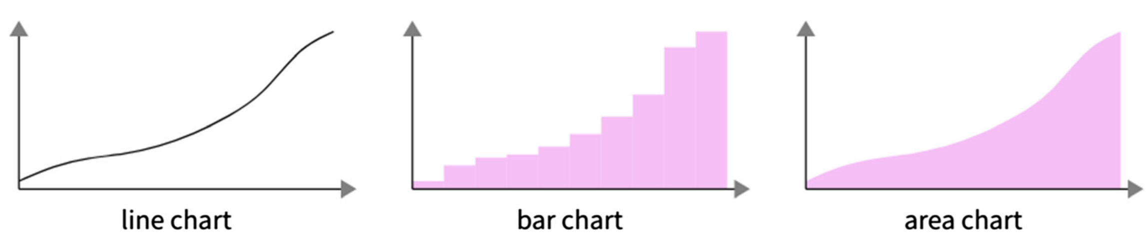


Fig. 5 The comparison among line chart, bar chart, and area chart. In each chart, the *x*-axis gives expression to the time, while the *y*-axis gives expression to the edge weight that represents the citation number

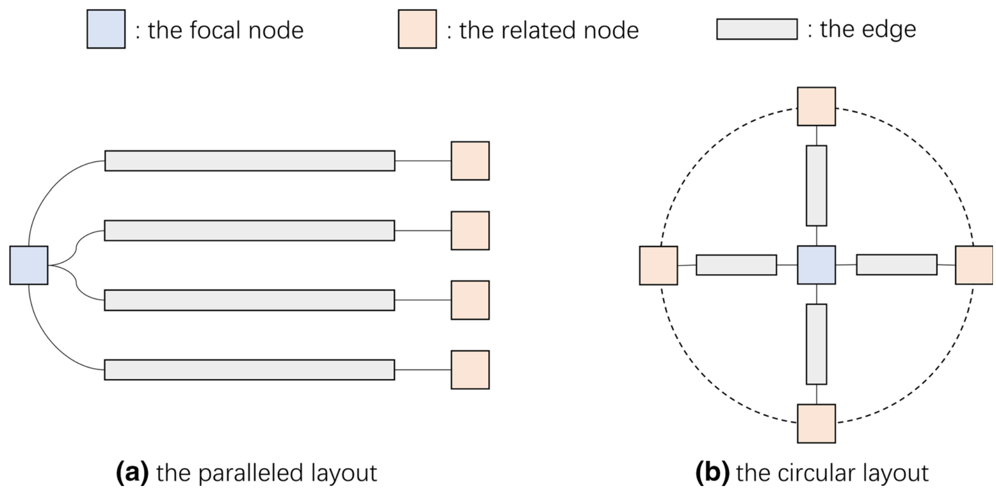


Fig. 6 The comparison between the paralleled layout and the circular layout

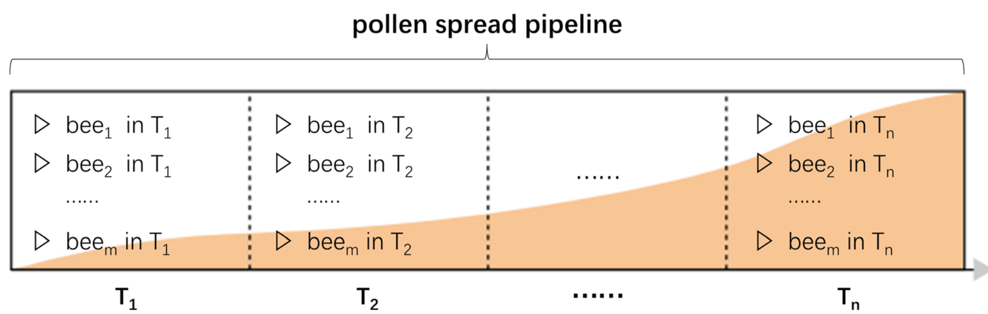


Fig. 7 The layout for the “bees.” The entire time is divided into *n* approximately average time spans. In each span, *m* “bees” are shown

visualized at the same time so that the details of these “flower terraces” can be shown. In the third level, a “flower terrace” is divided into multiple “flowers,” and the interface focuses on these “flowers” and another “flower terrace.” In the fourth-level interface which focuses on the influence of a specific “bee,” only the related “flower terraces” influenced by the specific “bee” are visualized, and the cooperative “bees” of the specific “bee” are visualized as well. For each analysis level, although the number of the “pollen spread pipelines” is specific, the “pollen spread pipelines” are always in the middle of the interface so that the consistency between different analysis levels can be kept.

4.4 Interaction

We design a series of interactions to change different states in the interface.

In order to change the current analysis level of the interface, a few interactions are necessary. Analysts can click the focal “flower terrace” to change the visual interface from the first level to the second level. The change from the second level to the third level can be done by clicking a “flower terrace.” And clicking a “bee” can change to the fourth level. What’s more, analysts can return to the previous level by clicking the blank area of the interface.

For the details in the “pollen spread pipelines,” we provide two buttons to change states. One button is designed to change the types of all “bees” in the interface, which has two options: “active” and “passive.” Another button is designed to change directions. This button has two options: “→” and “←,” which can change both directions of the “pollen stream” and the “bees” in each “pollen spread pipeline.”

In some cases, there are many related “flower terraces.” For the second-level interface where only several “flower terraces” are visualized at the same time, we provide the overview of all related “flower terraces” by showing cumulate rectangles with text information in the most right of the interface. A slider

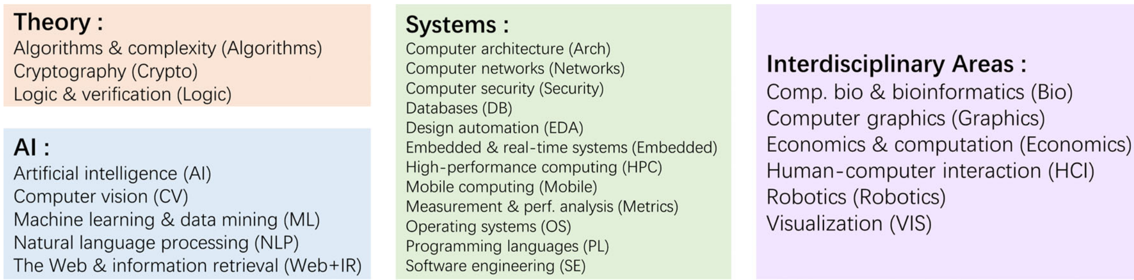


Fig. 8 The classification in CSRankings. The studies of Computer Science are classified into 4 categories and 26 fields. For the concise texts in visualization, we make abbreviations for all fields, which are shown as well

adjoins these rectangles. If analysts drag this slider, the related "flower terraces" and the "pollen spread pipelines" which are shown in the interface will vary, and the layout will remain unchanged.

5 Evaluation

In this section, we evaluate the effectiveness of our visual analysis system by reporting case studies and user study.

5.1 Data and data processing

We obtain the citation network data in Computer Science and do necessary processing.
Collect citation data. To collect the citation data in Computer Science, we use the information in CSRankings,[1] which classifies the studies of Computer Science into 4 categories and 26 fields (Fig. 8). The top conferences of each field are well selected in CSRankings. We obtain papers labeled by these top conferences from Microsoft Academic Graph (MAG) (Sinha et al 2015) which is an openly available academic database. The corresponding citations can also be obtained from MAG.

Aggregate collected papers. We aggregate the papers to form multiple sets by field classification in CSRankings. In this classification, virtual reality is considered as part of the visualization. To analyze visualization in a narrow concept, we treat virtual reality as an independent field. Based on these fields, we obtain 27 paper sets. To get more details, we further do a series of computations. For each set, we compute a few attributes, including the paper number, the citing number, and the cited number. Because the citation relationship is time-varying, we compute the annual citation numbers between the paper set of visualization and the paper sets of all related fields. Each paper has the corresponding authors who have promoting influence on the citation relationships. The authors can be classified as active authors and passive authors. Active authors are the authors who have cited other papers to publish their own papers, while passive authors are the authors whose papers have been cited by other papers. We divide all years into n year spans $[Y_1,...,Y_n]$. In a specific paper set S, an active author's influence in year y can be computed as the time t_y that he cited the papers in other sets to publish the papers in S. For each active author in a year span Y_i $(i \leq n)$, his influence is computed as $\sum_{\forall y \in Y_i} t_y$. We sort all active authors by their influence to get influential authors in Y_i. For passive authors, similar computation can be done.

Determine focal set. On account of the strong interest in the citation relationships between visualization and other fields, we choose the paper set of visualization as the focal set to perform the egocentric analysis. In CSRankings, IEEE VIS is selected as the top visualization conference. We get the set V_1 which consists of papers labeled with IEEE VIS in MAG and find that many papers are omitted. A possible reason is that many papers accepted to IEEE VIS are also published in IEEE Transactions on Visualization and Computer Graphics (TVCG) and most of these papers are labeled by TVCG rather than IEEE VIS in MAG. Using a collection of IEEE VIS papers (Isenberg et al 2017), we map titles in MAG to get another paper set V_2. For the final paper set of visualization, we get it by $V_1 \cup V_2$. After that, from the top conferences in CSRankings, we get the papers that have citation relationships with collected papers in IEEE VIS and classify them by different fields.

Par70 http://csrankings.org/.

Construct hierarchical structure. The hierarchical structure of the citation network is necessary. Each paper set based on CSRankings is at the first level. For obtaining the paper sets at the second level, we use the Field-of-Study (FoS) labels from MAG. In MAG, FoS labels represent scientific concepts, which form a six-level hierarchical structure (from L0 to L5) (Shen et al 2018). However, the hierarchical structures from L2 to L5 are automatically produced by an extended subsumption model and are not manually checked. What's more, because the mappings between FoS labels and papers are based on a machine learning algorithm rather than manual labeling, they are not completely correct. Using FoS labels directly to divide the paper sets cannot achieve an ideal result. Therefore, we do processing for the FoS labels. We reject the FoS labels at level 0 and level 1 as well as the FoS labels that represent the fields in CSRankings, and compute the top FoS labels in each paper set. We use these top FoS labels to do corresponding divisions for the paper set of the visualization field and the paper set of each field that has the citation relationship with the visualization field. At the second level, the annual citation numbers and authors can be computed in the same way.

Finally, we check all the data and find it relatively reasonable.

5.2 Case study 1: The temporal patterns in citation network

Based on the pollen-spread metaphors, we treat each first-level paper set as a "flower terrace" and each second-level paper set as a "flower." The "pollen spread pipelines" represent the citation relationships between different sets.

For the concise texts in the interface, we make abbreviations for all fields in CSRankings, which are shown in Fig. 8. Each category in CSRankings corresponds to a color scheme where the fields have different colors generated by ColorBrewer (Harrower and Brewer 2003). In the interface (Fig. 9), there are three "scutcheons" which show different attributes of each field: "N," "→," "←." "N" represents the number of the papers that took part in the citation relationship. For the "flower terrace" on the left, "→" represents the number that the papers of this field were cited by the papers of other fields, while "←" represents the number that the papers of this field cited the papers of other fields. For the "flower terrace" on the right, the meanings of "→" and "←" are opposite to those for the "flower terrace" on the left. The current state of the button "pollen direction" above means that the relationship direction is bidirectional now. The "bees" are not analyzed in this case.

Firstly, we get an overview of the citation network (R1). The visualization field has the citation relationships with almost all fields in Computer Science, which means that the cross-field characteristic of visualization is typical. Then, we begin to explore the attributes of each field (R2). It is apparent that the citing number is more than the cited number in the visualization field (Fig. 9a). And the top three fields with the most papers related to visualization are human–computer interaction, computer graphics, and computer vision (Fig. 9b). The interface shows that Cryptography, Economics & computation, Logic & verification have the least papers related to visualization (Fig. 9c). In the future, the papers of these fields can be considered to cite more papers of visualization, and the papers of visualization can be considered to cite more papers of these fields. For the dynamic citation numbers in the middle of the interface (Fig. 9d), we find that the citation numbers at two citation directions generally increase each year (R3).

When we click the "flower terrace" that represents the visualization field, the interface changes to show the citation relationships that the visualization papers were cited by the papers of each related field (R5). In order to explore the opposite relationships, we click the button "pollen direction" to change the direction. In the past few years, the number that the visualization papers cited the papers of human–computer interaction has increased rapidly (Fig. 10a). We notice that computer graphics had a different tendency compared to others (Fig. 10b). In the visualization field, the interest level of computer graphics papers has changed during the development. The number that the visualization papers cited the papers of computer graphics increased to a peak in about 2002 and decreased continuously after that. We try to give a possible interpretation of this phenomenon. In the initial stage of the visualization field, most papers focused on scientific visualization that had close relationship with computer graphics, which probably caused the increase in the number that the visualization papers cited the papers of computer graphics. With the booming development of information visualization and visual analysis, the number of papers about scientific visualization gradually decreased, which probably caused the decrease in the number that the visualization papers cited the papers of computer graphics.

In order to learn more about the citation relationship that the visualization papers cited the papers of computer graphics, we click the "flower terrace" that represents computer graphics. The interface shows the

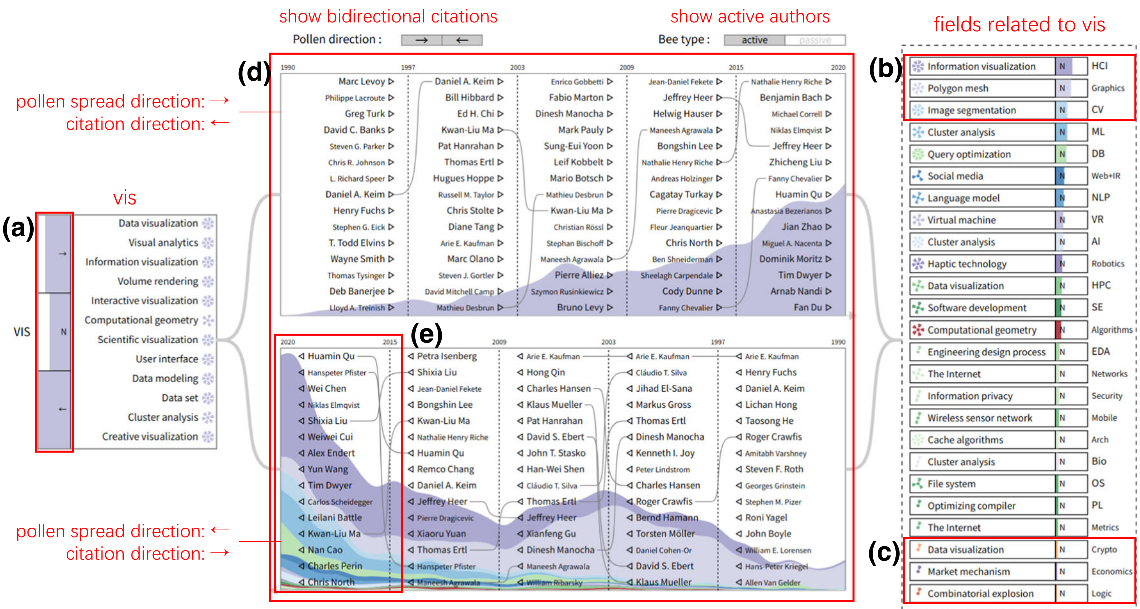


Fig. 9 The interface for the citation relationship between visualization and all related fields in Computer Science. From the "flower terrace" on the left, we can analyze the attributes of the visualization field (**a**). From the "flower terraces" on the right, we can analyze the fields that have the most papers related to visualization (**b**) and the fields that have the least papers related to visualization (**c**). In the middle of the interface, we can analyze the dynamic citation numbers at two citation directions (**d**). And we notice that an increasing number of Chinese researchers cited the papers of other fields to publish visualization papers in recent years (**e**)

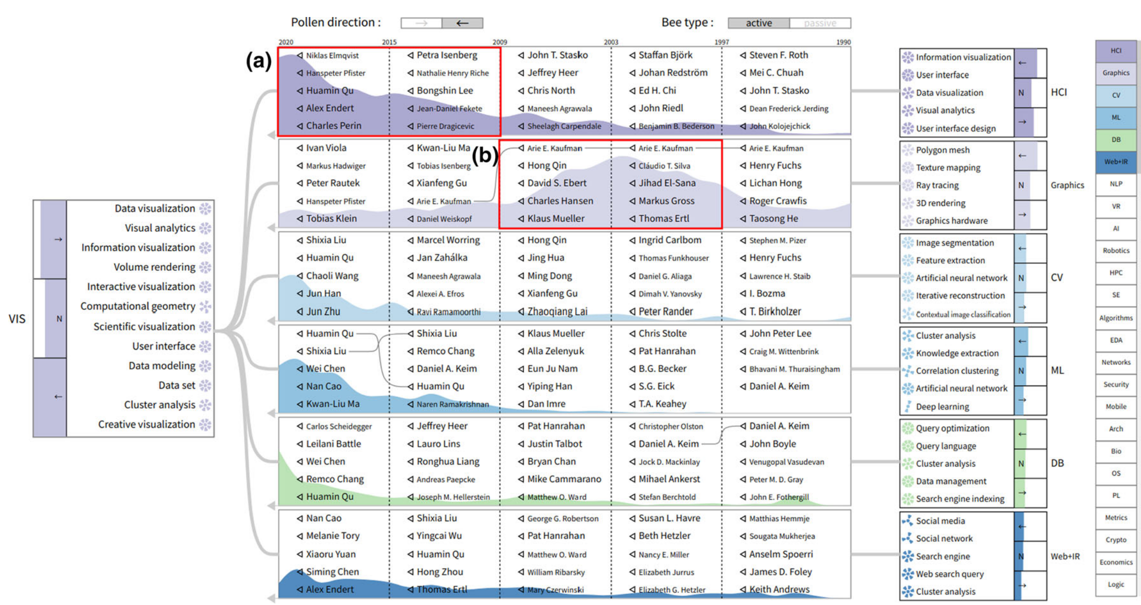


Fig. 10 The interface that shows how the papers of each related field were cited by the visualization papers. The number that the visualization papers cited the papers of human–computer interaction increased rapidly for the past few years (**a**). The number that the visualization papers cited the papers of computer graphics increased to a peak in about 2002 and decreased continuously after that (**b**)

citation relationships that the visualization papers cited the papers of the top FoS labels in computer graphics (R5). Each FoS label represents a scientific concept. The main scientific concepts that visualization cited from computer graphics include polygon mesh, texture mapping, and so on (Fig. 11a). For almost all the FoS labels shown in the interface, the number that the visualization papers cited the papers of them generally

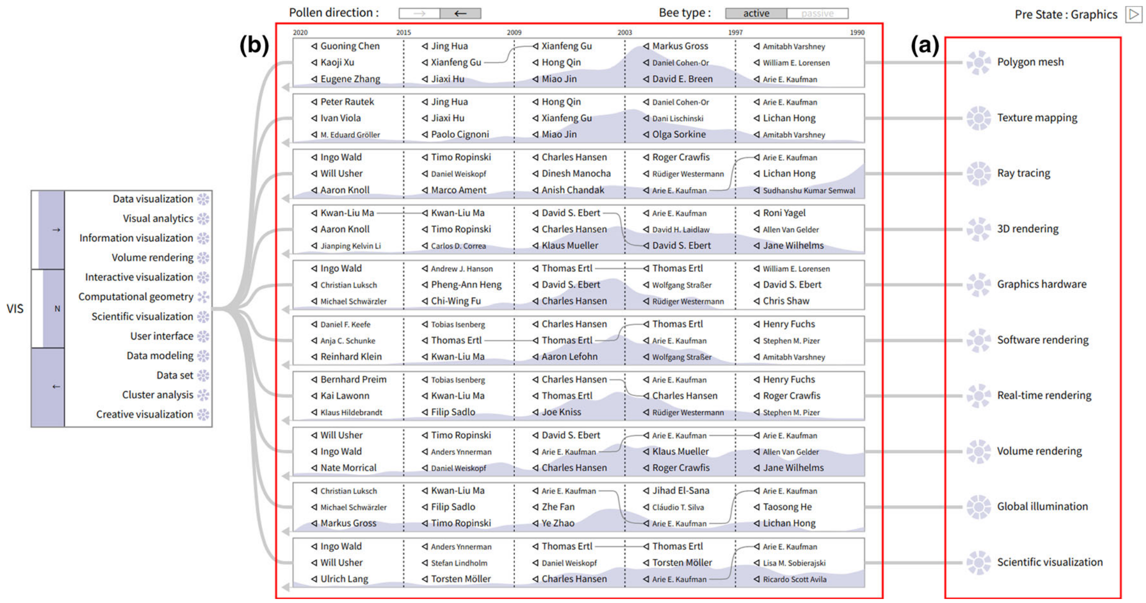


Fig. 11 The interface that shows how the visualization papers cited the papers of the top FoS labels in computer graphics. On the right of the interface, the top FoS labels in computer graphics are shown (**a**). In the middle of the interface, the number that the visualization papers cited the papers of the FoS labels generally decreased in recent years (**b**)

decreased in recent years (Fig. 11b), which means that the decrease in the interest level from the visualization field reflected in most of scientific concepts about computer graphics rather than a small amount of scientific concepts.

Besides, we also explore the citation relationships that the visualization papers cited the papers of the top FoS labels in machine learning & data mining (R5). We focus on cluster analysis and deep learning which we are particularly interested in (Fig. 12). The papers about cluster analysis were continuously popular to the visualization papers, which is probably because cluster analysis can effectively provide technical support in many visualization papers. With the rapid development of deep learning in recent years, the relevant papers have become more and more popular to the visualization papers.

5.3 Case study 2: The promoting influence of authors

The authors have promoting influence on the citation relationships, which can be treated as "bees" based on the pollen-spread metaphors. Our visual analysis system can explore their influence as well. The triangles that represent authors are visualized in each "pollen spread pipeline."

We begin to explore the authors in the interface, which shows the citation relationship between visualization and all related fields in Computer Science (Fig. 9). The current state of the button "bee type" means that the active authors are shown. We glance over the authors shown in the interface and notice that an increasing number of Chinese researchers cited the papers of other fields to publish visualization papers in recent years (Fig. 9e) (R4).

We hope to learn more about how a specific author influenced the citation relationships. We select the "bee" that represents Huamin Qu and explore how he cited the papers of other fields to publish visualization papers (Fig. 13) (R5). We find that he has cited papers of more than ten fields in Computer Science,

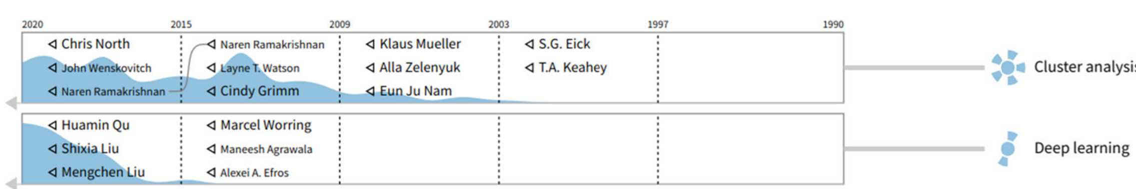


Fig. 12 The partial interface that shows how the visualization papers cited the papers of cluster analysis and deep learning

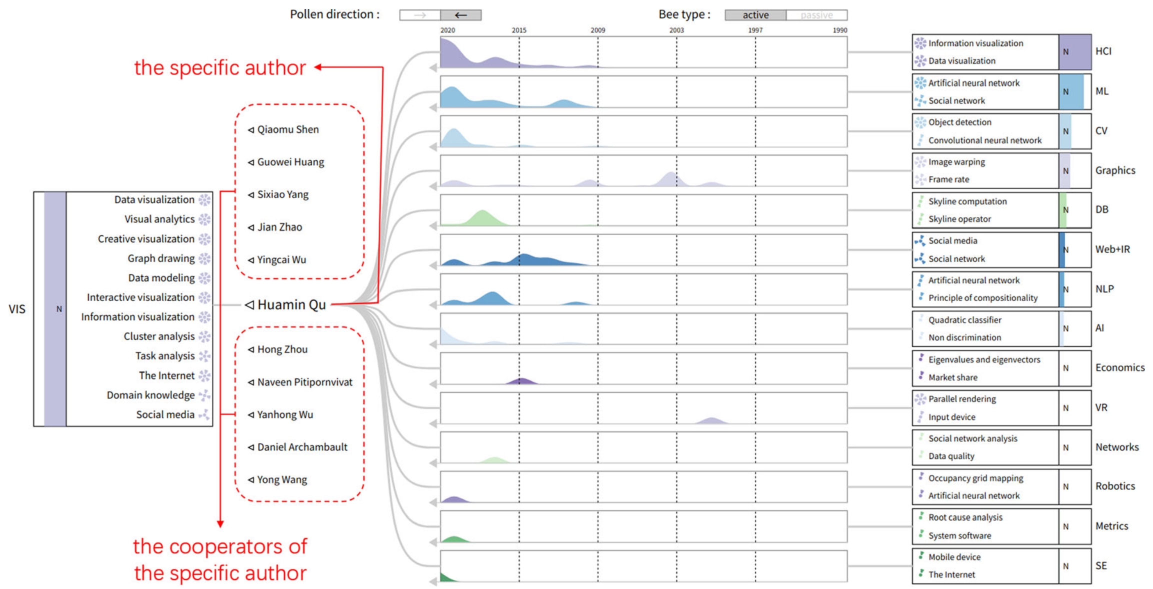


Fig. 13 The interface that shows how Huamin Qu cited the papers of other fields to publish papers in the visualization field

including human-computer interaction, machine learning & data mining, and so on. At the beginning of these citations, he mainly cited a few papers of computer graphics and virtual reality. Gradually, he cited the papers of more and more fields. From the "flowers" of each "flower terrace" on the right, we learn the main scientific concepts about the papers that he has cited.

In order to explore the passive authors, we click the button "bee type" to change the type of authors. Figure 14 shows the influential passive authors whose papers published in other fields were cited by the visualization papers (R4). Many curves connect the same "bees" which appear in two sequential year spans. This means that the visualization field persistently cited the papers of a few authors for a long time.

From the passive authors in another citation direction, we select Jeffrey Heer (Fig. 15a) and John T. Stasko (Fig. 15b) to analyze how their visualization papers were cited by the papers of other fields (R5). Their papers both attracted citations from the papers of more than ten fields and were popular to the papers of human–computer interaction.

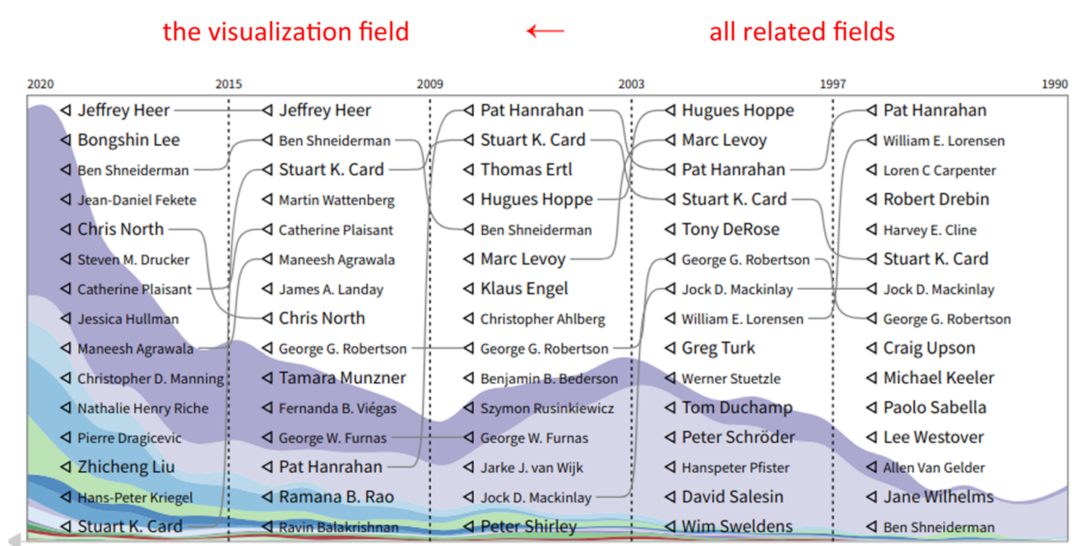
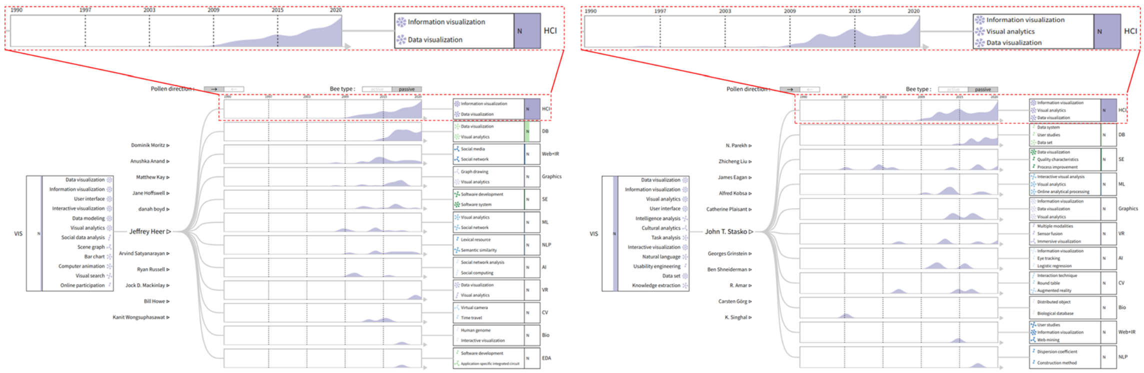


Fig. 14 The passive authors in the citation relationship that visualization cited all related fields. Their papers published in other fields were cited by the visualization papers. Many curves connect the same "bees" that appear in two sequential year spans, which means that the visualization field persistently cited the papers of a few authors for a long time



(a) Jeffrey Heer **(b)** John T. Stasko

Fig. 15 The interfaces that show how the visualization papers of Jeffrey Heer (**a**) and John T. Stasko (**b**) were, respectively, cited by the papers of other fields. Their papers were popular to the papers of human–computer interaction

5.4 User study

We conducted user study to validate the design of our system, including the volunteer experiment and the expert interview. In the user study, our system was still used to analyze the citation relationship between the visualization field and other fields in Computer Science.

Volunteer experiment. We recruited fifteen students from the same university to take part in the experiment. Eight of them are graduate students that research visualization, while others are graduate students that research other fields. We started with a tutorial to explain the purpose and design of our visual analysis system in 10 min. Then, each volunteer was asked to complete several tasks described in Table 1. These tasks involve different analysis aspects of our system. T1-T4 are straightforward, while T5-T6 are exploratory. The volunteers were asked to take a screenshot of the corresponding content after completing each task so that we could measure their accuracy. Moreover, we also measured their completion time for each task.

After all volunteers had completed the exploration using our visual analysis system, we analyzed the corresponding results. The accuracy and average completion time of each task are shown in Fig. 16. All volunteers could accurately complete T1-T2, while the accuracy of T3-T6 were, respectively, 86.7%, 93.3%, 86.7%, and 73.3%. We talked to the volunteers to find out the possible reasons why some of them could not complete tasks correctly. For T3, some volunteers forgot that the citation direction is the opposite direction of "pollen spread," so they chose the wrong citation direction to analyze. For T4, a volunteer did not change the current state of the authors by interaction, so he still analyzed the active authors. For T5 and T6, the main reason is similar to that of T3, and some volunteers stated that they could not understand the meanings of active authors and passive authors clearly.

Expert interview. We interviewed the three visualization experts mentioned in the requirement analysis again. After the experts became familiar with our system and did enough exploration freely, we asked them to comment on our system. We summarized the comments as follows. All experts were impressed by the design of our system. They affirmed the effectiveness of the visual metaphors and the egocentric analysis. An expert stated, "I like the design of pollen spread pipeline. It is ingenious that the dynamic citation numbers and the authors are shown together in each pipeline. I could find abundant information based on

Table 1 The tasks in the volunteer experiment

Requirement	Task
Overview (R1)	T1: Find all related fields that have the citation relationships with the visualization field
Node analysis (R2)	T2: Find the top three related fields that have the closest citation relationships with the visualization field
Edge analysis (R3)	T3: Conclude the trend of the citation relationship that the papers of all related fields cited the visualization papers
Promoter analysis (R4)	T4: Find the influential passive authors in each of the bidirectional citation relationships between the visualization field and all related fields in 1990-1997
Multiple-level (R5)	T5: Explore the citation relationship that the visualization papers cited the papers of computer graphics
	T6: Explore how Huamin Qu cited the papers of other fields to publish the visualization papers

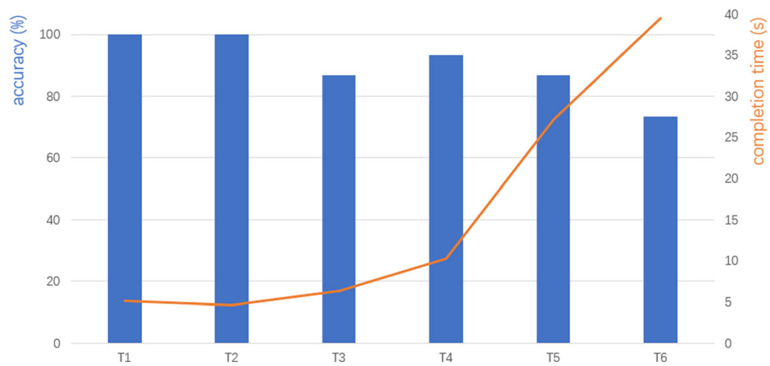


Fig. 16 The accuracy and average completion time of T1-T6 in the volunteer experiment

this design. For example, the number that visualization cited human-computer interaction had an extremely rapid increase in recent years, and Daniel A. Keim influenced both of the bidirectional citations between visualization and other fields." Another expert who studies visualization and machine learning stated, "The system enables me to do exploration effectively. I found that the visualization papers cited many papers about clustering and the citation number increased to a peak in about 2013. I also found that Chinese visualization researchers have cited more papers of machine learning compared to foreign researchers in recent years. The representative researchers are Huamin Qu and Shixia Liu." In addition, the experts gave some suggestions for our system. "The papers should be analyzed to explain and verify the obtained findings and inferences." "The meanings of active authors and passive authors are not easy to understand. Different visual designs may make sense." "More interactions should be designed to analyze the influence of authors." Based on these suggestions, we will make great efforts to improve our system in the future.

Overall, we gained positive results from the user study, which can evaluate the effectiveness of our visual analysis system.

6 Discussion

Limitations. Our visual analysis system has several limitations. Firstly, when designing visualizations based on the pollen-spread metaphors, we treat the edges as "pollen spread pipelines" that lack entity mapping in the real world, which may make analysts feel confused. Secondly, the number of the related fields that are shown once in our system is fixed, which causes the difficulty of exploring more related fields at the same time and comparing any two related fields, particularly when there are many related fields in total. Thirdly, the visual encoding for "flower" only shows the time span of the papers in each subfield, which cannot provide more information to analysts. Lastly, our system emphasizes egocentric analysis and uses the paralleled layout shown in Fig. 6 to visualize the focal node individually on the left, which causes the low data-ink ratio in the visual interface.

Generalization. In this paper, we focus on the egocentric visualization of citation network. Additionally, our visual analysis system can be used to analyze the networks which have multivariate nodes and time-varying edges after simple modification. A typical example is the reposting network in social media. In the reposting network, each node represents a social media user, while each edge represents the relationship that a user reposts the message initially posted by another user. Each message has a topic. All topics can be considered to promote the reposting relationship, which can be treated as promoters in the network. Because each user represented by a node has multiple attributes, all nodes can form hierarchical node sets based on different attributes. As a result, the pollen-spread metaphors can correspond to different components of the reposting network. After determining the focal set and doing necessary computations, our system can be used to explore how the different user sets at different levels repost the messages from each other dynamically and how the topics influence the reposting relationships.

7 Conclusions

In this paper, we design and implement a novel egocentric visual analysis system of citation network. We construct a new hierarchical egocentric network to perform multiple-level analysis for the citation relationships between a specific field and other fields. Based on the pollen-spread metaphors which can interpret the constructed network impressively, rich visualizations and interactions are designed to obtain the temporal patterns of the citation network. The effectiveness of our visual analysis system has been evaluated through case studies and user study.

Acknowledgements This work was supported by National Natural Science Foundation of China (Large scale literature data visual analysis based on science map, No.62172295).

References

Beck F, Burch M, Diehl S et al (2017) A taxonomy and survey of dynamic graph visualization. Comput Graph Forum 36(1):133–159. https://doi.org/10.1111/cgf.12791

Bryan C, Ma K, Fu Y (2013) An interactive visualization interface for studying egocentric, categorical, contact diary datasets. In: Rokne JG, Faloutsos C (eds) Advances in social networks analysis and mining 2013, ASONAM '13, Canada - August 25 - 29, 2013. ACM, pp 771–778. 10.1145/2492517.2492636

Burch M, Diehl S (2008) Timeradartrees: visualizing dynamic compound digraphs. Comput Graph Forum 27(3):823–830. https://doi.org/10.1111/j.1467-8659.2008.01213.x

Byron L, Wattenberg M (2008) Stacked graphs - geometry & aesthetics. IEEE Trans Vis Comput Graph 14(6):1245–1252. https://doi.org/10.1109/TVCG.2008.166

Cao N, Lin Y, Sun X et al (2012) Whisper: tracing the spatiotemporal process of information diffusion in real time. IEEE Trans Vis Comput Graph 18(12):2649–2658

Cao N, Lin Y, Du F et al (2016) Episogram: visual summarization of egocentric social interactions. IEEE Comput Graph App 36(5):72–81. https://doi.org/10.1109/MCG.2015.73

Chen S, Chen S, Wang Z et al (2019) D-map+: interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. ACM Trans Intell Syst Technol 10(1):11:1-11:26. https://doi.org/10.1145/3183347

Chen W, Xia J, Wang X et al (2019) Relationlines: visual reasoning of egocentric relations from heterogeneous urban data. ACM Trans Intell Syst Technol 10(1):2:1-2:21. https://doi.org/10.1145/3200766

Crnovrsanin T, Shilpika Chandrasegaran SK et al (2021) Staged animation strategies for online dynamic networks. IEEE Trans Vis Comput Graph 27(2):539–549. https://doi.org/10.1109/TVCG.2020.3030385

Cui W, Wang X, Liu S, et al (2014) Let it flow: a static method for exploring dynamic graphs. In: Fujishiro I, Brandes U, Hagen H, et al (eds) IEEE pacific visualization symposium, PacificVis 2014, Yokohama, March 4-7, 2014. IEEE computer society, pp 121–128. 10.1109/PacificVis.2014.48

Erten C, Harding PJ, Kobourov SG, et al (2003) Graphael: graph animations with evolving layouts. In: Liotta G (ed) Graph Drawing, 11th International symposium, GD 2003, Perugia, September 21-24, 2003, revised papers, lecture notes in computer science, vol 2912. Springer, pp 98–110. https://doi.org/10.1007/978-3-540-24595-7_9

Greilich M, Burch M, Diehl S (2009) Visualizing the evolution of compound digraphs with timearctrees. Comput Graph Forum 28(3):975–982. https://doi.org/10.1111/j.1467-8659.2009.01451.x

Harrower M, Brewer CA (2003) Colorbrewer.org: an online tool for selecting colour schemes for maps. Cartogr J 40(1):27–37

He Q, Zhu M, Lu B, et al (2016) Mena: Visual analysis of multivariate egocentric network evolution. In: 2016 International conference on virtual reality and visualization (ICVRV), pp 488–496. https://doi.org/10.1109/ICVRV.2016.88

Heimerl F, Han Q, Koch S et al (2016) Citerivers: visual analytics of citation patterns. IEEE Trans Vis Comput Graph 22(1):190–199. https://doi.org/10.1109/TVCG.2015.2467621

Herr BW, Duhon RJ, Börner K, et al (2008) 113 years of physical review: Using flow maps to show temporal and topical citation patterns. In: 12th International conference on information visualisation, IV 2008, 8-11 July 2008, London, IEEE computer society, pp 421–426. https://doi.org/10.1109/IV.2008.97

Isenberg P, Heimerl F, Koch S et al (2017) Vispubdata.org: a metadata collection about IEEE visualization (VIS) publications. IEEE Trans Vis Comput Graph 23(9):2199–2206. https://doi.org/10.1109/TVCG.2016.2615308

Jiang X, Zhang J (2016) A text visualization method for cross-domain research topic mining. J Vis 19(3):561–576. https://doi.org/10.1007/s12650-015-0323-9

Law P, Wu Y, Basole RC (2018) Segue: overviewing evolution patterns of egocentric networks by interactive construction of spatial layouts. In: Chang R, Qu H, Schreck T (eds) 13th IEEE conference on visual analytics science and technology, IEEE VAST 2018, Berlin, October 21-26, 2018. IEEE, pp 72–83. https://doi.org/10.1109/VAST.2018.8802415

Li Q, Shen Q, Ming Y, et al (2017) A visual analytics approach for understanding egocentric intimacy network evolution and impact propagation in mmorpgs. In: Weiskopf D, Wu Y, Dwyer T (eds) 2017 IEEE pacific visualization symposium, PacificVis 2017, Seoul, April 18-21, 2017. IEEE computer society, pp 31–40. https://doi.org/10.1109/PACIFICVIS.2017.8031576

Li Z, Zhang C, Jia S et al (2020) Galex: exploring the evolution and intersection of disciplines. IEEE Trans Vis Comput Graph 26(1):1182–1192. https://doi.org/10.1109/TVCG.2019.2934667

Liu Q, Hu Y, Shi L, et al (2015) Egonetcloud: Event-based egocentric dynamic network visualization. In: Chen M, Andrienko GL (eds) 10th IEEE conference on visual analytics science and technology, IEEE VAST 2015, Chicago, October 25-30, 2015. IEEE computer society, pp 65–72. https://doi.org/10.1109/VAST.2015.7347632

Nhon DT, Pendar N, Forbes AG (2016) Timearcs: visualizing fluctuations in dynamic networks. Comput Graph Forum 35(3):61–69. https://doi.org/10.1111/cgf.12882

Peng D, Tian W, Lu B, et al (2018) Dmnevis: a novel visual approach to explore evolution of dynamic multivariate network. In: IEEE international conference on systems, man, and cybernetics, SMC 2018, Miyazaki, October 7-10, 2018. IEEE, pp 4304–4311. https://doi.org/10.1109/SMC.2018.00728

Rufiange S, McGuffin MJ (2013) Diffani: visualizing dynamic graphs with a hybrid of difference maps and animation. IEEE Trans Vis Comput Graph 19(12):2556–2565. https://doi.org/10.1109/TVCG.2013.149

Shen Z, Ma H, Wang K (2018) A web-scale system for scientific knowledge exploration. In: Liu F, Solorio T (eds) Proceedings of ACL 2018, Melbourne, July 15-20, 2018, system demonstrations. Association for computational linguistics, pp 87–92. https://doi.org/10.18653/v1/P18-4015

Shi L, Wang C, Wen Z et al (2015) 1.5d egocentric dynamic network visualization. IEEE Trans Vis Comput Graph 21(5):624–637. https://doi.org/10.1109/TVCG.2014.2383380

Shin M, Soen A, Readshaw BT, et al (2019) Influence flowers of academic entities. In: Chang R, Keim DA, Maciejewski R (eds) 14th IEEE Conference on visual analytics science and technology, IEEE VAST 2019, Vancouver, October 20-25, 2019. IEEE, pp 1–10. https://doi.org/10.1109/VAST47406.2019.8986934

Sinha A, Shen Z, Song Y, et al (2015) An overview of microsoft academic service (MAS) and applications. In: Gangemi A, Leonardi S, Panconesi A (eds) Proceedings of the 24th international conference on world wide web companion, WWW 2015, Florence, May 18-22, 2015 - Companion Volume. ACM, pp 243–246. https://doi.org/10.1145/2740908.2742839

Steinböck D, Gröller ME, Waldner M (2018) Casual visual exploration of large bipartite graphs using hierarchical aggregation and filtering. In: 2018 International symposium on big data visual and immersive analytics, BDVA 2018, Konstanz, October 17-19, 2018. IEEE, pp 1–10. https://doi.org/10.1109/BDVA.2018.8533894

Sun M, Zhao J, Wu H et al (2019) The effect of edge bundling and seriation on sensemaking of biclusters in bipartite graphs. IEEE Trans Vis Comput Graph 25(10):2983–2998. https://doi.org/10.1109/TVCG.2018.2861397

Vehlow C, Beck F, Auwärter P et al (2015) Visualizing the evolution of communities in dynamic graphs. Comput Graph Forum 34(1):277–288. https://doi.org/10.1111/cgf.12512

Wu Y, Pitipornvivat N, Zhao J et al (2016) Egoslider: visual analysis of egocentric network evolution. IEEE Trans Vis Comput Graph 22(1):260–269. https://doi.org/10.1109/TVCG.2015.2468151

Yoon T, Han H, Ha H, et al (2020) A conference paper exploring system based on citing motivation and topic. In: 2020 IEEE Pacific visualization symposium, PacificVis 2020, Tianjin, June 3-5, 2020. IEEE, pp 231–235. https://doi.org/10.1109/PacificVis48177.2020.1010

Zhao J, Glueck M, Chevalier F, et al (2016) Egocentric analysis of dynamic networks with egolines. In: Kaye J, Druin A, Lampe C, et al (eds) Proceedings of the 2016 CHI conference on human factors in computing systems, San Jose, May 7-12, 2016. ACM, pp 5003–5014. https://doi.org/10.1145/2858036.2858488