

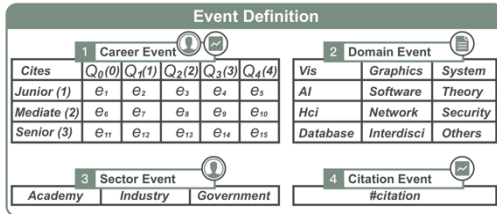
Supplementary Material

- [Data Formats](#)
- [Implementation Details](#)

Data Formats

Categories	Questions	Data Formats
(1) Raw Data (Section 3.2 Data Description)	Career Data	<p>Description</p> <ul style="list-style-type: none"> • <i>Background:</i> we collect both LinkedIn data and other sources (researchers' personal websites and their institutional webpages) into the same career format. • <i>Data unit:</i> the data unit is a researcher. • For every researcher, we collected the education and working experiences. • For each type of experience, there is a list of events recording information including: <ul style="list-style-type: none"> • <code>Event: <working period, job title, organization, location></code>
	Bibliographic Data	<p>Description</p> <ul style="list-style-type: none"> • <i>Background:</i> Aminer holds almost all data in the VIS field. • <i>Data unit:</i> the data unit is a paper. We further arrange the data into researcher-based. • Attributes we use in a paper include: <ul style="list-style-type: none"> • <code><year, author_list, venue_name></code>
	Citation Data	<p>Description</p> <ul style="list-style-type: none"> • <i>Background:</i> we used Google Scholar API to get the citation data by year. • <i>Data unit:</i> the data unit is a researcher. • For each researcher, the numbers of citations are listed by year. <ul style="list-style-type: none"> • <code><year, number_of_citations></code>

(2) Sequential Data
(Section 4.1 Data Preprocessing)



Career Sequences

Description

- Each researcher is a sequence.
- Each event in the sequence is defined by both title and citation ranks (the left figure)
 - `<year, career_event_type>`

Sector Sequences

Description

- Each researcher is a sequence.
- Each event in the sequence is one of “academy”, “industry”, and “government agency”.
 - `<year, sector_event_type>`

Domain Sequences

Description

- Each researcher is a sequence.
- Each event in the sequence is a list of number of papers in each research domain (12 categories in total)
 - `<year, domain_frequency_list>`
 - `domain_frequency_list: [#paper_in_category1, ..., #paper_in_category12]`

Citation Sequences

Description

- Each researcher is a sequence.
- Each event in the sequence is the number of citations in a year

Implementation Details

Implementation Categories	Steps	Remark
(1) Data Collecting (Section 3.2 Data Description)	1-1. Extracted names of 1200+ VIS researchers.	By using DBLP data to apply the constraints: <ul style="list-style-type: none"> those who have published more than two TVCG papers in which the largest time gap is more than five years as potential VIS researchers.
	1-2. Manually checked these names and filtered out those from other fields (about 90 researchers) and finally obtained over 1,100 VIS researchers.	We manually checked names based on researchers' profiles, such as publications and homepages.
	1-3. Searched for three data sources based on researchers' names.	We use researchers' other information (e.g., institutions) to distinguish those with the same names manually.
(2) Data Preprocessing (Section 4.1 Data Preprocessing)	2-1. Career data: semi-automatically tagged the job titles and sectors.	We conduct the tagging in four steps (take job title for an example): <ol style="list-style-type: none"> Maintained a table to classify subjective job titles (e.g., assistant professor) into different categories in career data, which was generated based on domain knowledge. Used this table to classify more job titles For those new titles that were not included in the table, we manually checked them, either adding them into the table or just labeling the specific title with a category. Used both job title categories and the researchers' tenure in academic research to determine the title

		<p>ranks (i.e., junior, intermediate, and senior).</p> <p>Sector tagging was conducted in a similar way.</p>
	<p>2-2. Bibliographic data: semiautomatically tagged the papers into different research domains based on an existing classification (Please refer to our paper <i>Section 4.1 Data Preprocessing</i>)</p>	<p>We took similar steps as the tagging of job titles above:</p> <ol style="list-style-type: none"> 1. Some main venues were already classified in the existing classification. We did not need to do more. 2. For those venues not in the classification table, we also maintained a table to classify them into different domain categories using keywords semi-automatically.
	<p>2-3. Bibliographic data: extracted the dynamic ego-networks of a researcher by year using paper author list.</p>	--
	<p>2-4. Citation data: classify them into five categories using Quartile (Please refer to our paper <i>Section 4.1 Data Preprocessing</i>).</p>	--
	<p>2-5. Construct four sequences based on the results of 2-1 ~ 2-4. Please refer to our paper <i>Section 4.1 Data Preprocessing</i> (the second paragraph) and Fig. 2-B for more details.</p>	--
<p>(3) MIA framework (<i>Section 4.3 Multi-factor Impact Analysis (MIA)</i>)</p>	<p>Please refer to <i>Section 4 Data Analysis</i> for more details.</p>	--