SpreadLine: Visualizing Egocentric Dynamic Influence

Supplementary Materials

Yun-Hsin Kuo, Dongyu Liu, and Kwan-Liu Ma

A Supplementary Information for Sect. 4 Designing SpreadLine

A.1 Examples of different optimization focuses

We present the comparison of two optimization focuses that SpreadLine offers: **vertical space optimization** and **straight line optimization**.

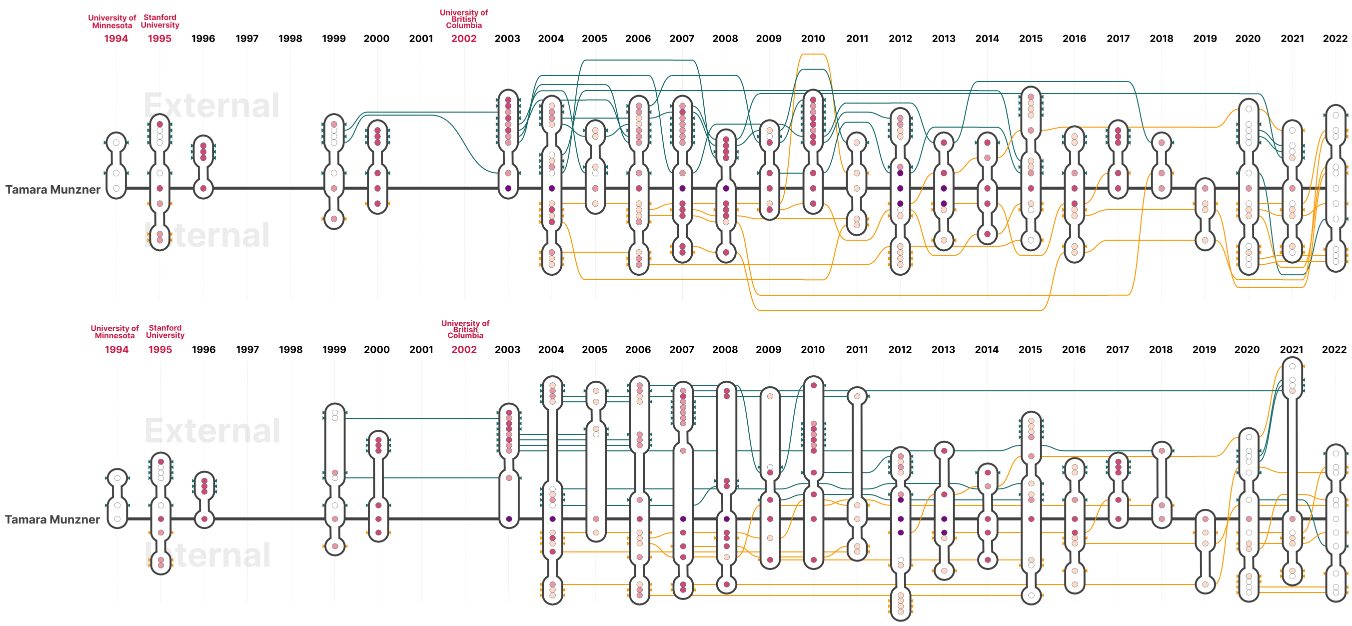

Figure A.1: Career evolution of Tamara Munzner in two different optimization focuses. Top: vertical space optimization. Bottom: straight line optimization.

A.2 Algorithmic logic of compact stage

We briefly describe the heuristics logic in these two optimization focuses. At this stage, the order of non-idle entities is already determined in the previous two stages. In the general logic of our algorithm, it always iterates through non-idle entities in a chronological, top-down order. Once their positions are assigned, the algorithm subsequently positions idle entities.

Vertical space optimization. For non-idle entities at each timestamp, we ensure the consecutive entities always maintain a distance based on their identity (i.e., in primary or secondary blocks) and configurations (e.g., stacking the same category). We assign the heights of the idle entities once the ego is aligned to the same height across the timestamps. For each idle entity, as each idle timestamp is always between two non-idle timestamps, we consider two cases: (1) this entity is constantly on one side of the ego; (2) this entity will cross the ego line at some point. In the first case, we attempt to arrange this idle entity to share the same height at either of the two non-idle timestamps to minimize the line wiggles. If this attempt is not viable (e.g., the assignment is taken) or it is the second case, we assign it to the spot that is closest to the current assignment on the same side.

Straight line optimization. Similar to vertical space optimization, we first align the ego to obtain initial heights for all non-idle entities. Next, for each non-idle entity, we consider all of its initial heights on the same side and reassign it to the minimum (if it is above the ego) or maximum (if below) of these heights. However, we keep the initial assignment if the reassignment is already occupied or is too far from the initial assignment (a minimum distance of at least 10 entities). The reassignment will also reposition other entities by shifting them to maintain the ordering of entities. Much vertical space in SpreadLine representations is the result of this design rationale. For idle entities, while they could be easily repositioned to the reassignments at their non-idle timestamps, we check whether other entities (both idle and non-idle) would be affected based on this simple update. The scenarios generally include: (1) All entities are unaffected, so the update is executed; (2) Some entities are affected but not reassigned yet, so we shift them and execute the update; and (3) The update is not feasible, so we reassign the idle entities in the same way as in vertical space optimization.

A.3 Design considerations

We have discussed 4 major design design considerations in the manuscript. In the following, we presented two more topics.

Line continuity. The line-based design allows SpreadLine to portray the dynamics of entities with continuous lines across timestamps. The block distinction and the points visually cut each line into fragmented lines, resulting in the increased difficulty of tracking the dynamics of entities. However, both point mark and block representation serve different and important analysis purposes, the fragmented lines are a consequent tradeoff we chose to accept. SpreadLine supports line hovering as one user interaction to address this limitation. When the user hovers over an alter line, other lines (excluding the ego line) would be de-highlighted, along with the blocks where the hovered alter is absent. This allows users to efficiently identify the time points at which the hovered alter interacted with the ego, thereby understanding its dynamics.

Straightened ego line and space division. During the early development of SpreadLine, we found it difficult to locate the ego across timestamps when the ego line was not straightened. Therefore, we decided to straighten the ego line to facilitate better visual tracking of the ego and to focus on its dynamic relationships with alters, which becomes a crucial design element in SpreadLine. This rationale leads to two divided spatial compartments, resulting in our current design of space division. Consequently, an inherent limitation is that space division can only show two kinds of information at a time, which is a tradeoff we accepted due to the egocentric design of SpreadLine. As space regions draw consumers' attention the most, we intend space division to provide a high-level characterization of entities. If more than 2 relationships need to be presented, we recommend utilizing other encodings to present such information, such as line colors, node colors, or block distinction.

B Supplementary information for Sect. 5 Using SpreadLine

We provide the link to our source code: `https://anonymous.4open.science/r/SpreadLine-08AF`. The layout generation is implemented in Python, and the rendering is in vanilla JavaScript with d3.js. Our plans involve full support for SpreadLine in Typescript.

```python
SpreadLiner = SpreadLine()
SpreadLiner.load(network, config={
    'source': 'Source',
    'target': 'Target',
    'time': 'Time',
    'weight': 'weight',
})
SpreadLiner.load(entities, config={
    'entity': 'entity',
    'color': 'color',
}, key='line')
SpreadLiner.load(f'{path}/color.csv', config={
    'time': 'date',
    'entity': 'entity',
    'context': 'attitude',
}, key='node')
layout = pd.read_csv(f'{path}/layout.csv')
SpreadLiner.load(layout, config={
    'timestamp': 'time',
    'id': 'name',
    'posX': 'posX',
    'posY': 'posY',
}, key='content')

SpreadLiner.center(ego='Danny Masterson')
SpreadLiner.configure({'squeezeSameCategory': True,
                       'bandStretch': [['2023-09-07', '2023-09-10'], ['2023-09-17', '2023-09-19']],
                       'minimize': 'line'})
result = SpreadLiner.fit(width = 2600, height = 500)
```

Figure B.1: The interface of how SpreadLine framework users provide data and configure hyperparameters.

SpreadLine framework has 4 main methods: load, center, configure, and fit. We briefly describe the purpose of each method.

Load. This method takes the data in the form of Pandas DataFrames, json objects, or file paths. For SpreadLine to know which column stores what information, SpreadLine requires "config", which is a dictionary object of the fixed keys referring to the actual column names. "key" is also required for SpreadLine to know which processing should be performed on what kind of data.

Center. This method takes the assigned ego and prepares the data correspondingly, e.g., egocentric network construction.

Configure. This method allows framework users to control all the hyperparameters described in Sect. 5.2. For example, "minimize" refers to the optimization focus specification.

Fit. Calling this method will perform the optimization and compute the layout that fits the specified size if provided. It returns all rendering results to produce SpreadLine representations.

C Supplementary information for Three Case Studies

For each case study, we provide the data structure used in SpreadLine and corresponding prompts for large language models.

C.1 Usage Scenario: Animal Disease Outbreak

Table C.1 shows the data structure that SpreadLine takes in. No data transformation is applied to be used in SpreadLine. This usage scenario presents 38 transaction records among 9 farms, including 46 entries of their health conditions. Due to data sensitivity, we are acquiring permission to share the anonymized data.

In addition, we provide distance information from Farm SI (ego) for the findings discussed in the manuscript. The distance from Farm SI to Farm FCT is 37.49 miles (60.33 kilometers); to Farm FI is 3.54 miles (5.69 kilometers), which is within a viable range for airborne transmission; to Farm FAO is 14.16 miles (22.79 kilometers), and to Farm FAN is 12.48 miles (20.08 kilometers).

Task ID	Task Description	Date	Source	Target	Quantity
xxxx	...	2020-02-28	FA	FB	xxxx
xxxx	...	2020-03-13	FB	FC	xx
xxxx	...	2020-03-07	FA	FC	xxx

(a) Trading records among farms.

Entity	Date	Context
FA	2020-03-18	0.5
FB	2020-03-16	1.0
FC	2020-03-11	0.641

(b) Unhealthy status of farms.

Entity	Type	Latitude	Longitude
FA	sow	xx.xxxxxx	xx.xxxxxx
FB	sow	xx.xxxxxx	xx.xxxxxxx
FC	finish	xx.xxxxxxx	xx.xxxxxx

(c) Contextual information of farms.

Table C.1: Data structures of animal disease outbreak used in SpreadLine. (a) The dynamic directed network topology, where quantity is treated as the edge weight; (b) The node color in SpreadLine representations; (c) The farm type refers to the line color, where the coordinates are used in contextual affinity view.

C.2 Case 1: Public Reaction to #MeToo Event

Table C.2 shows the data structure used in SpreadLine. Due to the policy of X, we can disclose only the tweet ID but not the tweet content. As described in section 6.1, we have collected 48 tweets and employed ChatGPT to perform data processing. This results in 56 entities and 184 relations in total, but the egocentric network of Danny Masterson only involves 49 entities and 178 relations, as reported in section 6.1.

C.2.1 Data Structure

Source	Target	Context	Date	ID
Danny Masterson	Sentence	received	2023-09-07	1699988429134848337, ...
Sentence	30 years to life	quantified as	2023-09-07	1699988429134848337, ...
Ashton Kutcher	Danny Masterson	supports	2023-09-08	1700367726315704773, ...

(a) Knowledge graphs of tweets.

ID	Stance
1699988429134848337	Oppose
1699892072957051118	Support
1699907781657313630	Support

(b) Stance of tweets.

Entity	Category
Scientology	Identity
Defence	Fact
Justice	Opinion

(c) Categories of entities.

Entity	Date	Stance	Support	Neutral	Oppose	Count	Positive	Neutral	Negative
Scientology	2023-09-07	Support	1	0	0	1	0	1	0
Justice	2023-09-07	Support	3	0	0	3	2	0	1
Crusaders	2023-09-08	Oppose	0	1	0	1	0	0	1

(d) Contextual information of entities.

Table C.2: Data structure of #Metoo incident used in SpreadLine. (a) The relations from the knowledge graphs of tweets; (b) The space division in SpreadLine representations; (c) The line color in SpreadLine representations; (d) The stance distribution of entities, along with its sentiment.

C.2.2 ChatGPT Prompts

Below we describe the prompts we provide to GPT-4 through the interface. We choose ChatGPT-4 over other models as it has access to the Internet, so we may consider its knowledge as ground truth when performing these tasks.

Task(1) Annotate the stance for the sentencing of the ego.

Given this tweet, determine what is the poster's stance for the sentencing of Danny Masterson. You can only choose between support, neutral, or oppose.

Task(2) Construct the knowledge graph for a given post.

Construct the knowledge graph of this tweet. Print the edges in a .csv file

Task(3) Determine the poster's attitude towards an entity.

You will get an entity and a tweet. Determine what is the speaker's attitude towards the entity in this tweet. You can only choose between positive, negative, and neutral.

Task(4) Annotate the entity.

You will be given an entity. Based on your knowledge about Danny Masterson's #meToo case, annotate it as one of the following: (1) objective fact; (2) subjective opinion; (3) specific individual; (4) generic group

C.3 Case 2: Career Evolution of VIS Researcher

In this case study, we use the DBLP citation network dataset v14, collected by AMiner. It contains more than 5 million papers recorded on DBLP. For the original data structure, please refer to their documentation. Here, we report the attributes used in this case study: "authors", "year", "n_citation", and "keywords". Note that "authors" stores a list of author instances, including their name, id, and organizations. Table C.3 shows the data structure used in SpreadLine. Note that we omit the ID as it is the unique identifier used in AMiner.

For the Munzner case, we extracted 242 relations with 123 co-authors from her 97 papers. For the Heer case, we extracted 365 relations with 198 co-authors from his 133 papers.

C.3.1 Data Structure

Source	Target	Year	Paper ID	Citation
Ben Shneiderman	Catherine Plaisant	1989	xxx	56
Miriah D. Meyer	Tamara Munzner	2010	yyy	97
Miriah D. Meyer	Hanspeter Pfister	2010	yyy	97

(a) Relations of authors.

Name	Year	Citation	Affiliation
Ben Shneiderman	1973	103	State University of New York at Stony Brook...
Tamara Munzner	2009	34	University of British Columbia, Canada
Tamara Munzner	2013	186	University of British Columbia

(b) Details of authors.

Entity	Year	posX	posY
Tamara Munzner	1994	0.36272	0.42020
Tamara Munzner	1995	0.47924	0.33280
Computer graphics	1994	0.33657	0.43123

(c) Contextual information of authors.

Table C.3: Data structure of VIS researchers used in SpreadLine. (a) The network constructed from the research papers; (b) Author information used in node color and space division in SpreadLine; (c) The normalized word embeddings, used in contextual affinity view.

C.3.2 ChatGPT Prompts

In this case study, we use OpenAI API to obtain the word embeddings for each keyword. We ask ChatGPT-3.5 to explain a given keyword, and then access the word embedding of the explanation

```
    messages = [{
            "role": "system",
            "content": """You are a research assistant.
                The user needs you to explain some terminologies. """
            },{
            "role": "user",
            "content": "What does {} mean?".format(entity)
            }]
```

D SUPPLEMENTARY INFORMATION FOR USABILITY STUDY

D.1 Participant Background

Table D.1 provides self-reported background information about the participants. In the following, we describe the summarized context of the participants' familiarity with both topics. Note that all participants indicated that they had never seen or used this specific academic publication dataset used in our work.

Expert participants' network visualization experiences primarily stem from research projects; their familiarity with academic publication datasets is from reviewing other publications or using similar datasets in their research.

Beginner participants have generally used node-link diagrams in their course projects to present information. One participant participated in a user study involving academic publication datasets, while the others suggested that they were only aware of the types of information such datasets might contain.

For **novice participants**, N2 and N3 have seen node-link diagrams in publications and coursework. N2 also explained, "I have a rough idea about publication datasets because my siblings have worked on this before, but I have never used them."

	Visualization Experience	Expertise	Network Visualization Experience (1-7)	Publication Dataset Experience (1-7)
E1	6 years	Computer graphics	5	4
E2	4 years	Geospatial visualization	4	3
E3	6 years	Visual analytics	4	4
E4	3 years	Text visualization	6	6
B1	2 years	Computer graphics	4	4
B2	2 years	HCI and information visualization	4	3
B3	8 months	Information visualization	4	3
N1	None	Statistics	1	1
N2	None	Energy systems	2	2
N3	None	Computer architecture	2	1

Table D.1: Self-reported participant profiles. The last two columns use a 7-point Likert-scale. 1: I don't know what they are; 4: moderate knowledge; 7: extremely knowledgeable.

D.2 Usability and Utility Questions

As we have provided the quantitative and qualitative questions in the manuscript, here we list out the statements for usability and utility.

Usability assessment.

1. SpreadLine is easy to learn.
2. SpreadLine is easy to interpret.
3. SpreadLine helps visually connect four aspects of the network.
4. SpreadLine is able to reveal informative insights about the ego.
5. SpreadLine is not frustrating.

Utility assessment.

1. SpreadLine helped identify global patterns (e.g., collaboration tendency).
2. SpreadLine helped discover global pattern changes (e.g., number of co-authors).
3. SpreadLine helped identify other co-authors of interest.
4. SpreadLine helped identify timepoints of interest.
5. SpreadLine helped understand a co-author's relationship dynamics with the ego

D.3 Static Images Used in the T1

Both Fig. D.1 and Fig. D.2 present the same information.

E TWO SUPPLEMENTARY SPREADLINE REPRESENTATIONS

In the following, we present two more researchers with larger collaboration networks. Note that we did not annotate the time labels in these representations.

E.1 Ben Shneiderman

The first researcher is Ben Shneiderman, who has published 281 papers with 498 co-authors from 1971 to 2022.
Figure D.1.1 shows the career evolution of Ben Shneiderman.
Our findings are as follows:

(1) Unlike the researchers presented in the manuscript, Ben mostly published by himself in the first 20 years of his career.

(2) He maintains a balanced level of external and internal collaboration throughout the timeline and appears to sustain long-term collaborations with many co-authors.

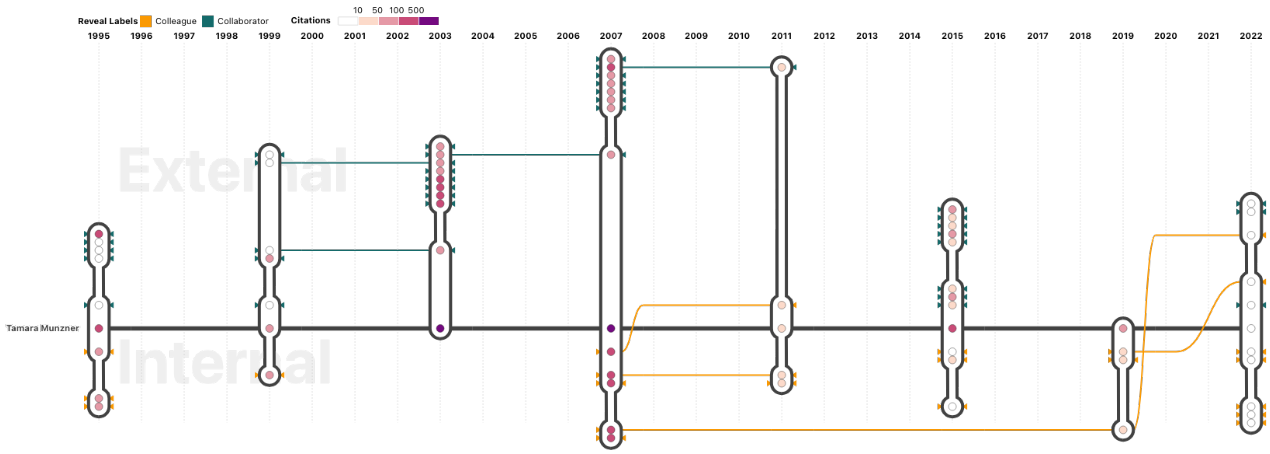

Figure D.1: The SpreadLine representation.

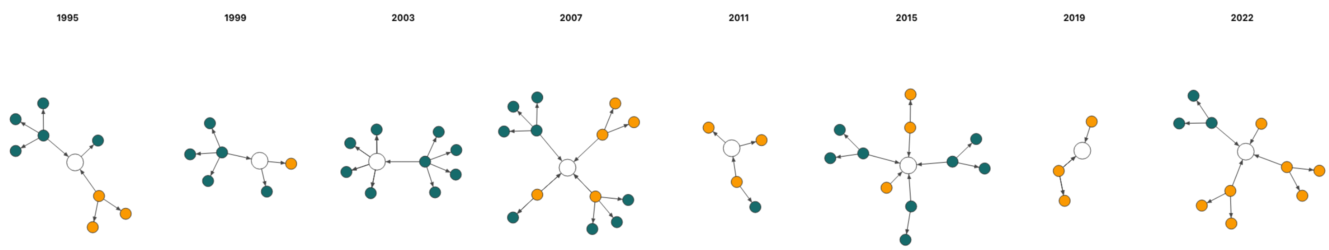

Figure D.2: The Node-link representation.

(3) In comparison to the researchers presented in the manuscript, Ben exhibits fewer line crossings over the ego line, indicating fewer organizational changes in his collaboration network.

We can further utilize the filtering interaction to identify his long-term collaborators, as shown in Figure D.1.2.

(4) We can observe that three co-authors, Catherine Plaisant, Terry Winograd, and Pattie Maes, are highlighted by the filter. However, both external collaborators didn't seem to maintain frequent collaboration with Ben, whereas Catherine consistently published papers with him almost every year as an internal collaborator.

E.2 Huamin Qu

The second researcher is Huamin Qu, who has published 214 papers with 368 co-authors from 1999 to 2022.
Figure D.2.1 shows the career evolution of Huamin Qu. Our findings are as follows:

(1) Huamin has maintained consistent internal collaboration since the beginning of his career.

(2) In recent years, although there appears to be a balanced mix of internal and external collaborations, the prevalence of yellow color on the external side indicates that many of his external collaborators were internal.

By filtering the crossings of the ego line, as depicted in Figure D.2.2, we can compare these two figures and observe that the majority of Huamin's long-term collaborators were internal collaborators.

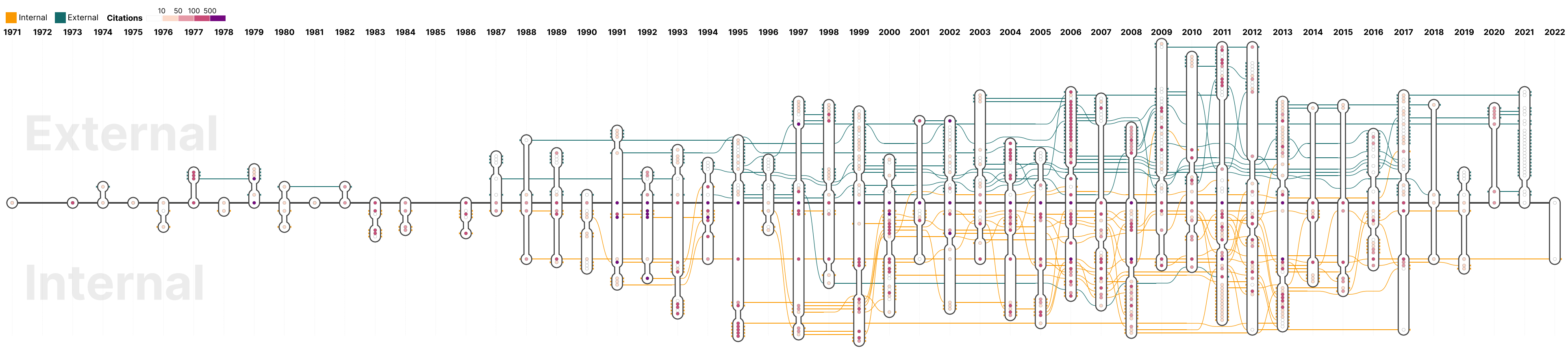

Figure D.1.1: Career evolution of Ben Shneiderman from 1971 to 2022.

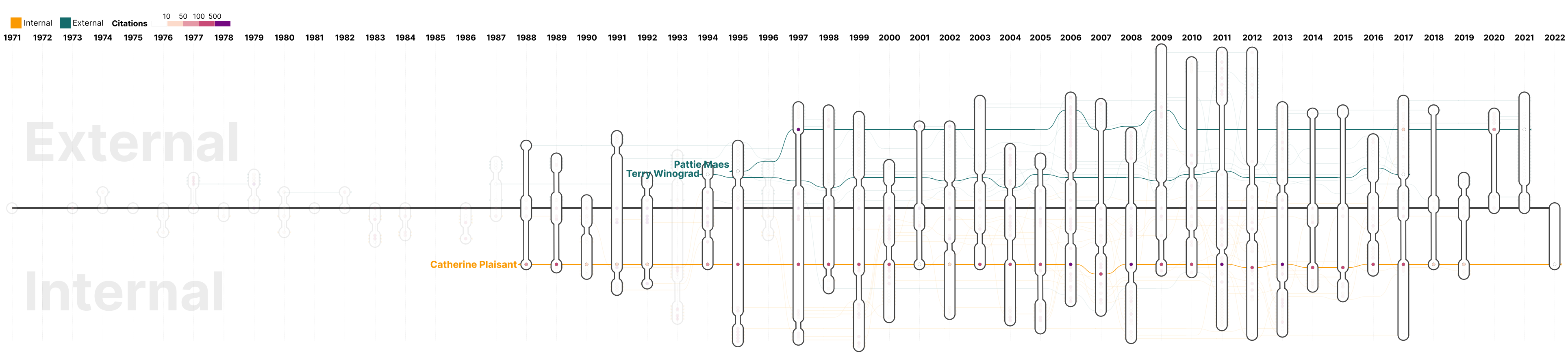

Figure D.1.2: Long-term collaborators of Ben Shneiderman over 20 years from 1971 to 2022.

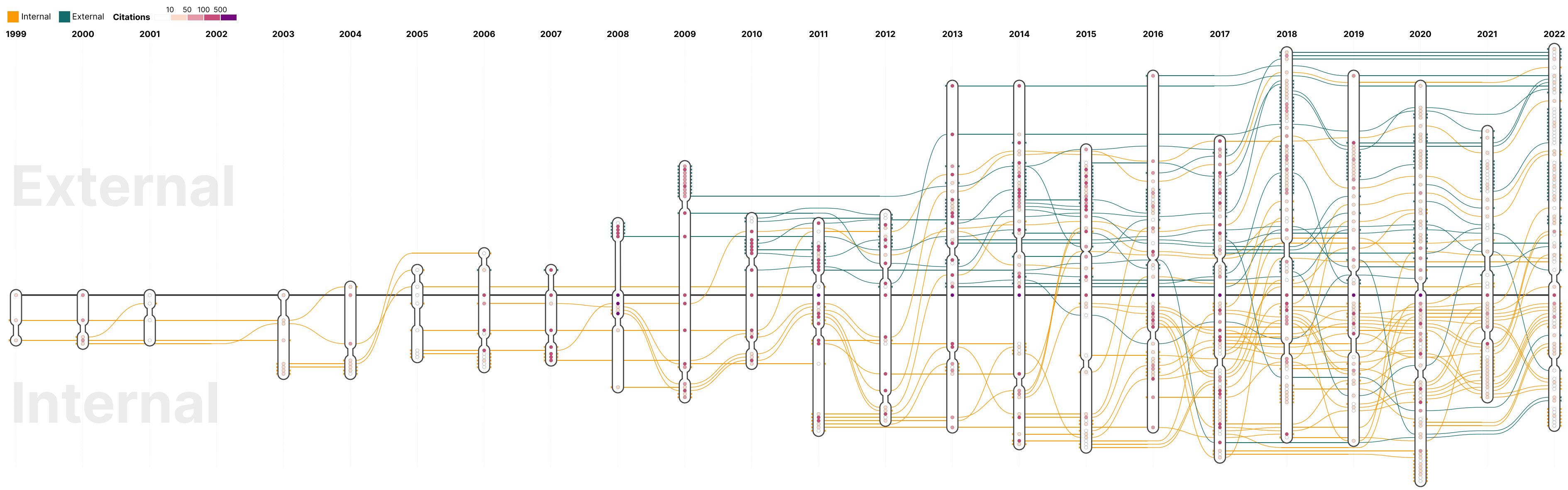

Figure D.2.1: Career evolution of Huamin Qu from 1999 to 2022.

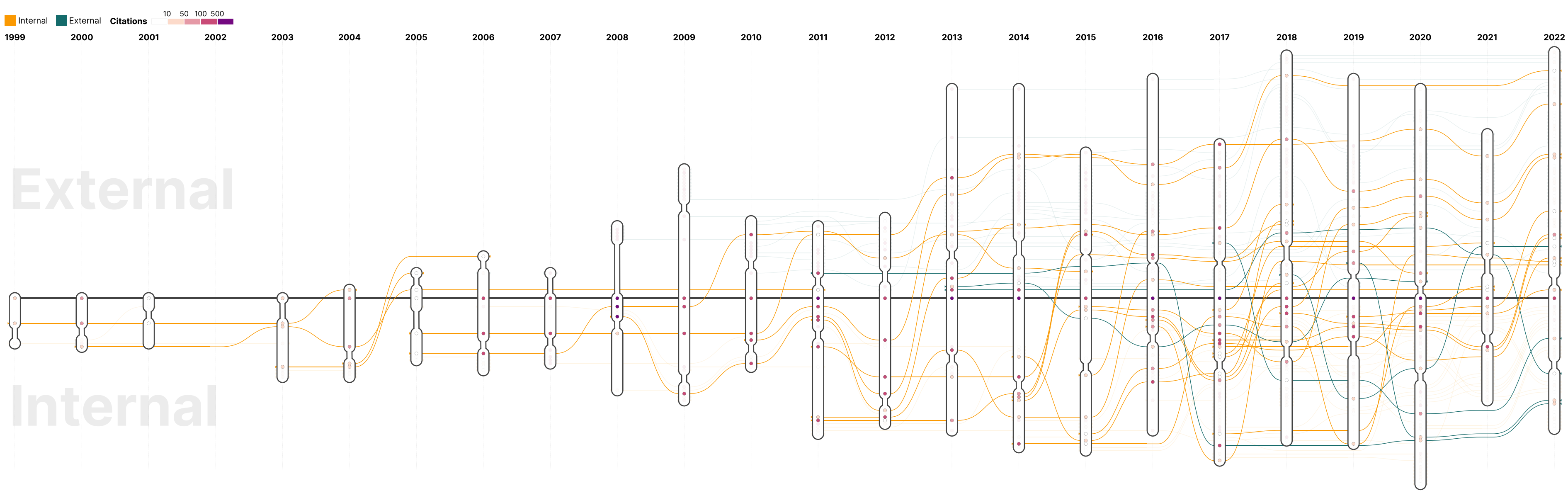

Figure D.2.2: Collaborators of Huamin Qu that had crossed the ego line from 1999 to 2022.