Extracting the Evolutionary Backbone of Scientific Domains: The Semantic Main Path Network Approach Based on Citation Context Analysis

Journal:	*Journal of the Association for Information Science and Technology*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Jiang, Xiaorui; Coventry University, Research Centre for Computational Sciences and Mathematical Modelling Liu, Junjun; Tsinghua University,
Keywords:	citation analysis < bibliometrics < informetrics < measurement < quantitative research < (research and analytic methods), scientometrics < measurement < quantitative research < (research and analytic methods), network analysis < data analysis < (research and analytic methods)

Extracting the Evolutionary Backbone of Scientific Domains: The Semantic Main Path Network Approach Based on Citation Context Analysis

Xiaorui Jiang[1], Junjun Liu[2]

[1] Coventry University, Coventry, UK, xiaorui.jiang@coventry.ac.uk

[2] ych1776@163.com

Abstract

Main path analysis is a popular method to extract the scientific backbone from the citation network of a research domain. Existing approaches ignored the semantic relationships between the citing and cited papers. This resulted in several adverse issues, such as long paths with incidental citations and poor semantic coherence among main path papers. This paper advocated the semantic main path network approach to alleviate these issues based on citation function analysis. A wide variety of SciBERT-based deep learning models were designed for identifying citation functions. Semantic citation networks were built by either including important citations, e.g., extension and motivation, or excluding incidental citations like background and future work. Semantic main path network was built by converging the top-*K* main paths extracted from various time slices of the semantic citation networks. Experiments on two research areas of computational linguistics demonstrated that, compared to semantics-agnostic approaches, the semantic main path network approach could discover novel topic branches and enrich branches with new important papers. Thus, it drew a more comprehensive and precise picture of domain evolution. In addition, the proposed approach could uncover more coherent development pathways between scientific ideas by identifying and preserving significant citations pertinent to scientific knowledge flow.

Keywords

Main path analysis; semantic main path network; Citation function classification; scientific backbone; knowledge flow

1. Introduction

There were many methods to extract the evolutionary pathways between the ideas of scientific publications based on analysing citation relationships, such as algorithmic historiography (Garfield et al., 2003), (knowledge flow through) semantic link networks (Zhuge, 2006), scientific historiograms (Lucio-Arias & Leydesdorff, 2008), and influence trajectories (Tu & Hsu, 2016). A related stream of studies first ranked the publications and then connected the selected important publications in some way (Bae et al., 2014; Zhang, et al., 2014; Tao et al., 2017; Ding et al., 2021). Amongst these, main path analysis (MPA), originally proposed in Hummon and Dorerian (1989), has become popular since Verspagen (2007). Verspagen (2007) laid out the algorithmic foundations of MPA. First, assign a weight to each citation edge to measure the number of knowledge diffusion paths, called *search paths*, from historical publications (source nodes) to recent publications (sink nodes). See Kuan (2020) for a discussion. Then, a main path of a citation network is a connected subnetwork of important citation edges connecting sources to sinks. In this sense, a main path is believed to roughly represent the development trajectories among the major ideas advancing the analysed scientific domain. By adding a pseudo-source (resp. pseudo-sink) connected to all sources (resp. sinks) of the reversed citation network, Batagelj (2003) proposed the efficient *search path counting* algorithm based on topological order. Based on this mathematical tool, several *main path extraction* methods were proposed, including *forward main path* (start from the highest weighted out-going edges from sources and iteratively explore subsequent out-going edges with the highest weights), *backward main path* (start from the highest weighted in-coming edges to sinks and iteratively explore the next in-coming edges with the highest weights) and *key-route main path* (start from the highest citation edge and explore both directions).

Most MPA methods did not take the semantic relationships between papers into consideration. They were citation semantics-agnostic. A direct consequence is semantically incoherent long main path caused by "incorrect" search path counting. Figure 1 illustrates the cause for this problem. In the right-top schematic image, the citation edges (A, B) and (B, C) are both neutral citations ("Neutral") while the citation edge (A, C) is an extension citation ("Extends"). Ignoring citation functions, the search path counts of (A, C) and (B, C) are larger than (A, C), so traditional approaches will add (A, C) and (B, C) to the main path. However, it is more reasonable to include the extension citation (A, C) in the main path. Some studies extended traditional MPA approaches to weighted citation networks by, for example, decaying longer paths in search path counting (Liu & Kuan, 2016), considering citation preferences according to discipline and publication time (Yu & Pan, 2021), or scaling search path count using citing paper's prestige (Yu & Sheng, 2021). However, these approaches did not perfectly solve the problem. For example, if B is highly cited, then Yu and Pan's approach will still choose (A, B) in main path exploration. There were also studies weighting citation edges based on similarities between citing and cited papers (Liu et al., 2014; Kim et al., 2022). However, topic similarity is less descriptive than paper authors' own citation motivation, aka citation function (Lyu et al., 2021).

We conjecture that semantic MPA using citation function analysis will generate better semantic coherence than traditional approaches. To illustrate, the left-most image in Figure 1 shows a vanilla main path network extracted using the traditional key-route MPA approach. The path length between A00-2018 and D07-1096 is very long (also refer to Sect. 4.2.1 for details): distance(A00-2018, D07-1096) = 16. The middle image shows the corresponding snapshot of the semantic main path network extracted by considering extension ("Ext") and motivation ("Mot") citations only. Now the distance is decreased to 5, which partially implies better semantic coherence of the main path from A00-2018 to D07-1096. For another example, the right-bottom image shows a snapshot of the semantic main path network extracted by further considering usage ("Use") and similarity ("Sim") citations. The path length between W96-0213 and W05-0516 is only 5, while the distance in the vanilla main path network is 17. To the best of our knowledge, this is the first paper which marries citation function classification to main path analysis. We proposed a systematic approach to semantic main path network extraction (Sect. 3) using citation function classification results (Sect. 2). Note that, there were also some recent studies relying on screening out incidental citations, aka citation importance classification, to perform knowledge flow analysis (Hassan et al., 2018) or scientific lineage extraction (Ghosh et al., 2021). Essentially, these approaches weighted citation edges by 1 (important) or 0 (incidental) and kept all non-incidental citation links without any further processing. Thus, this paper is methodologically different from them. Citation function classification provides us with more flexible ways to perform MPA. The superiority of the proposed approach was qualitatively justified using two case studies (Sect. 4).

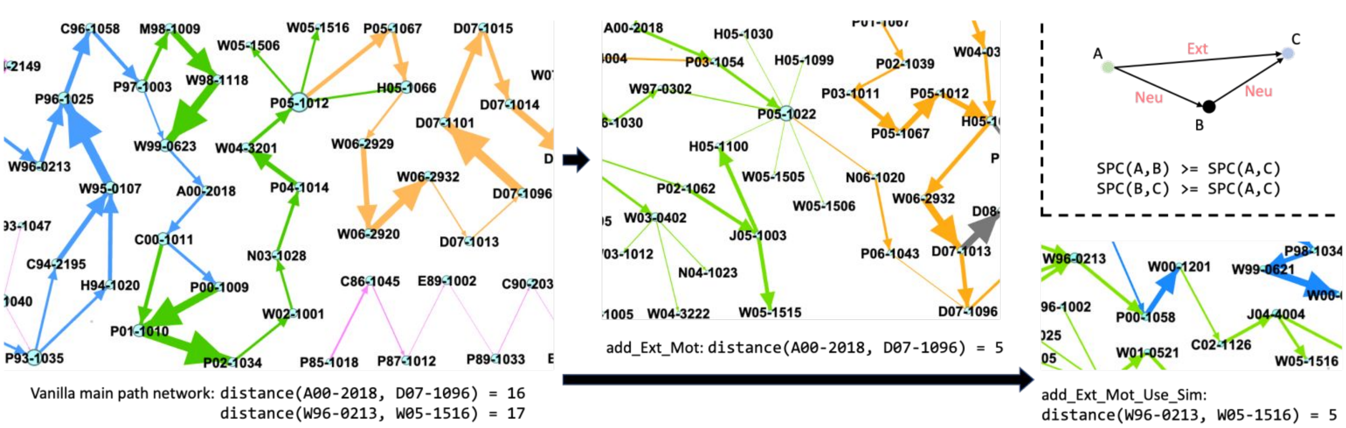

Figure 1. Motivations for Semantic Main Path Analysis (SPC: Search Path Count)

2. Citation Function Classification

2.1. Dataset and Annotation Schemes

We created a large citation function dataset by merging and re-annotating six existing datasets in the computational linguistics domain: `Teufel2010` (Teufel et al., 2006a; Teufel, 2010), `Dong2011` (Dong & Schäfer, 2011), `Jha2016` (Abu-Jbara et al., 2013; Jha et al., 2016), `Alvarez2017` (Hernández-Alvarez et al., 2017), `Jurgens2018` (Jurgens et al., 2018), and `Su2019` (Su et al., 2019). The source papers were crawled from ACL Anthology[1], ranging between early 1990s and late 2010s. Different annotation guidelines were adopted so all datasets must be re-annotated.

Here gives a brief description of the re-annotation effort (see Supplementary Sect. A for a detail account). All citation contexts were-reannotated according to Teufel et al.'s 12-class annotation scheme (Teufel et al., 2006a) plus a "Future" class about future work. The author and three postgraduate research students in natural language processing annotated and reached consensus on all instances. Note that, for the neutral function, we only included the "Neut" instances from `Teufel2010` and the "Neut" instances re-annotated from other categories of the six datasets. After re-annotation, consecutive citation strings were merged into a pseudoword "CITSEG" (meaning "citation segment") because in-text citations in the same citation segment must have the same function. For example, the citation sentence "SHRDLU (Winogard, 1973) was intended to address this problem." would be tokenized and rewritten to "["SHRDLU", "(", "CITSEG", ")", "was", "intended", "to", "address", "this", "problem", "."]". Some minority classes were still too small, so we merged "PModi" with "PBas" into "Basis", and re-annotated "CoCo-" into "CoCoGM" or "CoCoRes". This resulted in our own 11-class citation function scheme, which could be mapped to 7-class and 6-class schemes by category merging. Table 1 shows the statistics of the annotated set, named `Jiang2022`.

[1] https://aclanthology.org/

Table 1. Citation Function Scheme Mapping and CITSEG-level Statistics

Teufel2010+ (12+1 class)				Jiang2021 (11-class)			Jiang2021 (7-class)			Jurgens2018 (6-class)		
label	size		ratio	label	size	ratio	label	size	ratio	label	size	ratio
	citstr	citseg	citseg									
Future	97	85	2.21%	Future	85	2.21%	Future	85	2.21%	Future	85	2.21%
CoCoXY	200	152	3.94%	CoCoXY	152	3.94%	Background	1773	46.00%	Background	1615	41.90%
Neut	1924	1463	37.96%	Neutral	1463	37.96%*						
Weak	223	158	4.10%	Weakness	158	4.10%				ComOrCon	944	24.49%
CoCoGM	390	299	7.76%	CoCoGM	328	8.51%	ComOrCon	479	12.43%			
CoCo-	108	80	2.08%	CoCoRes	151	3.92%						
CoCoR0	107	100	2.59%									
PSup	123	100	2.59%	Support	100	2.59%	Similar	307	7.97%			
PSim	247	207	5.37%	Similar	207	5.37%						
PMot	365	288	7.47%	Motivation	288	7.47%	Motivation	288	7.47%	Motivation	288	7.47%
PUse	794	755	19.59%	Usage	755	19.59%	Uses	755	19.59%	Uses	755	19.59%
PModi	72	65	1.69%	Basis	167	4.33%	Extends	167	4.33%	Extends	167	4.33%
PBas	134	102	2.65%									
Total	4784	3854										

2.2. Citation Function Classification Models

We employed SciBERT to develop a series of deep learning models for in-text citation function classification. The citation context, currently fixed to 2 left and 3 right context sentences, were concatenated. Three feature representations were pooled from SciBERT encodings of the citation context. The CITSEG encoder extracted the encodings of the pseudoword "CITSEG" as the *citation representation* **h**. According to Lauscher et al. (2021), most citation instances' functions could be determined only using the citation sentence (abbr. citance). Therefore, the *citance representation* **s** was pooled from the citance words by the citance encoder. Similarly, the context encoder pooled the *context representation* **c** from the whole context. The final feature vector **f** was the concatenation of the three parts: **f** = [**h**; **s**; **c**], where the citation representation is mandatory because different citations in the same citance should have different feature representations, but citance and context representations were optional.

We tested two types of citation contexts. In a *sequential context*, no "[SEP]" (sequence separator) was inserted to separate context sentences. In this case, citance and context representations were directly pooled from the citance tokens and context tokens respectively. Two poolers were tested: max-pooling and self-attention (Munkhdalai et al., 2016). In a *hierarchical context*, "[SEP]" symbols were inserted after each context sentence. Then, context representation was pooled from the representations of its enclosed sentences and sentence pooler could take "[SEP]" as the third option in addition to max-pooling and self-attention.

There were in total 34 model variants[2]. Due to the large GPU time required to train all these model variants on different annotation schemes, we chose a subset to run initial experiments with the 11-class scheme and cherry-picked 11 relatively promising variants (Table 2). In Sect. 3.1, we will discuss how to analyse the per-class performances of different models and pick the appropriate models to perform semantic main path network analysis.

Table 2. Selected Citation Function Classification Models

ID	citseg_encoder (**h**)	context_type	sentence_pooler	citance encoder (**s**)	context_encoder (**c**)
1	O (used)	sequential	N/A	max_pool	max_pool
2-3	O	sequential	N/A	X (not used)	max_pool (2); self_attend (3)
4-6	O	sequential	N/A	max_pool (4); self_attend (5); X (6)	X
7-8	O	hierarchical	max_pool	X	max_pool (7); self_attend (8)
9-11	O	hierarchical	N/A	max_pool (9); self_attend (10); X (11)	X

3. Semantic Main Path Network Analysis

3.1. Model Selection: Precision or Recall

Per-class performances of the trained models showed that no single best model could beat others on all citation functions or on all annotation schemes (Supplementary Table S1-S3). Therefore, we decided to choose the most appropriate model and use it as a binary classifier to analyse a specific citation function. For example, the most pertinent citation function for main path network analysis seemed to be the extension class ("Basis" or "Extends"), meaning "ideational basis of" or "technical extension to" the cited work. The best extension model was selected from all 55 trained models (11 model variants by 5 seeds per model variant).

Figure 2 and 3 show the performances of a few top models for the extension and motivation classes respectively. The darker the colour, the higher the performance. We see that although the best model was model 4 (seed = 5171) with the 6-class scheme, its recall was relatively low. Considering the small size of the extension class, e.g. only 4.33% in our dataset, we decided to slightly **weigh recall over precision** and F1 **(recall-oriented)**. The final chosen model had a good F1 and the highest

[2] When **f** = **h**, depending on `context_type`, the number of model variants is 2. When **f** = [**h**; **s**], the number of model variants is: 2 (`context_type` = "sequential") + 2×3 (`context_type` = "hierarchical") = 8. When **f** = [**h**; **c**], if `context_type` = "sequential", the model variant number is 2; otherwise, if `context_type` = "hierarchical", it is $3 \times 2 = 6$ (3 sentence poolers by 2 context encoders). When **f** = [**h**; **s**; **c**], if `context_type` = "sequential", the model variant number is 2×2 (2 citance encoders multiplied by 2 context encoders) = 4; otherwise if `context_type` = "hierarchical", there are $2 \times 3 \times 2 = 12$ model variants (2 citance encoders by 3 sentence poolers by 2 context encoders). Therefore, there are in total 2 + 8 + (2 + 6) + (4 + 12) = 34 model variants.

recall, i.e. model 11 (seed = 47353, in solid red rectangle) trained with the 6-class scheme. Taking a similar recall-oriented approach, we chose model 7 (seed = 32491) trained with the 6-class scheme as the "best" motivation model.

We hoped that the semantic citation network could capture as many important citations as possible. Valenzuella et al. (2015) defined usage and extension as important citation relations. Therefore, we further added usage citations to the semantic citation network. According to Figure 4, although we still took a recall-oriented approach, we opted for model 7 (seed = 13249) trained with the 11-class scheme which achieved the highest F1, because the recall of the chosen model was already high enough and its precision was much higher than other candidates. To further enrich the semantic citation network, we also added similarity citations because Teufel's annotation guidelines say similarity is between problems and solutions rather than results. According to Figure 5, the selected model was model 11 (25603) trained with the 11-class scheme.

The other way is to delete incidental citations, e.g. neutral citations ("Neutral"/"Background") or future work citations ("Future") in our case. "Neutral" is the largest class, e.g. 37.96% in our dataset (Table 1). "Weakness" (4.10%) and "CoCoXY" (3.94%) could also be converted to "Background" (in the 7-class and 6-class schemes). Due to the dominating size of neutral citations and the relatively high performances on this class, we decided to **trade recall for precision (precision-oriented)** for the "Neutral"/"Background" and "Future" classes. According to Figure 6, model 2 (5171) with the 7-class scheme was chosen to recognise neutral citations ("Neutral"/"Background"). Note that this model subsumed weakness citations too (refer to Table 1). According to Figure 7, model 8 (32941) was used to screen out future work citations. The chosen model's 93.75% precision was high enough and its recall and F1 were both much better than other candidates.

Extends		Jiang2021 (11-class)					Jiang2021 (7-class)					Jurgens2018 (6-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
4	precision	70.37	65.52	63.33	62.50	74.07	65.52	76.00	63.64	70.97	64.71	80.77	56.76	52.94	62.50	68.97
	recall	55.88	55.88	55.88	58.82	58.82	55.88	55.88	41.18	64.71	64.71	61.76	61.76	52.94	58.82	58.82
	f1 score	62.30	60.32	59.38	60.61	65.57	60.32	64.41	50.00	67.69	64.71	70.00	59.15	52.94	60.61	63.49
8	precision	70.82	43.75	61.29	80.95	71.43	57.89	75.00	75.00	75.50	63.33	50.00	61.29	62.96	75.00	60.00
	recall	50.00	41.18	55.88	50.00	44.12	64.71	61.76	52.94	61.76	55.88	58.82	55.88	50.00	44.12	52.94
	f1 score	58.62	42.42	58.46	61.82	54.55	61.11	67.74	62.07	67.74	59.38	54.05	58.46	55.74	55.56	56.25
11	precision	53.85	61.29	70.83	55.88	63.33	72.00	64.71	61.09	57.58	57.89	65.52	50.00	72.00	53.85	67.65
	recall	61.76	55.88	50.00	55.88	55.88	52.94	64.71	55.88	55.88	64.71	55.88	41.18	52.94	61.76	67.65
	f1 score	57.53	58.46	58.62	55.88	59.38	61.02	64.71	58.46	56.72	61.11	60.32	45.16	61.02	57.53	67.65

Figure 2. Performances of Selected Best Models for the Extension Class

Motivation		Jiang2022 (11-class)					Jiang2022 (7-class)					Jurgens2018 (6-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
6	precision	57.14	60.27	63.79	56.72	60.66	59.32	67.21	62.90	71.15	56.06	62.96	58.33	60.94	55.07	53.52
	recall	68.97	75.86	63.79	65.52	63.79	60.34	70.69	67.24	63.79	63.79	56.82	60.34	67.24	65.52	65.52
	f1 score	62.50	67.18	63.79	60.80	62.18	59.83	68.91	65.00	67.27	59.68	60.71	59.32	63.93	59.84	58.91
7	precision	60.27	60.00	55.56	59.68	58.73	64.71	66.67	62.96	62.50	55.74	56.42	54.22	57.35	68.25	60.78
	recall	75.86	62.07	60.34	63.79	63.79	56.90	65.52	58.62	68.97	59.62	67.24	77.59	67.24	74.14	53.45
	f1 score	67.18	61.02	57.85	61.67	61.16	60.55	66.09	60.71	65.57	57.14	61.42	63.83	61.90	71.07	56.88
11	precision	62.90	60.00	66.67	68.85	54.41	66.67	67.27	58.57	57.14	58.90	54.41	66.13	69.84	61.29	74.55
	recall	67.24	62.07	68.97	72.41	63.79	68.97	63.79	70.69	68.97	74.14	63.79	70.69	75.86	65.52	70.69
	f1 score	65.00	61.02	67.80	70.59	58.73	67.80	65.49	64.06	62.50	65.65	58.73	68.33	72.73	63.33	72.57

Figure 3. Performances of Selected Best Models for the Motivation Class

Usage		Jiang2022 (11-class)					Jiang2022 (7-class)					Jurgens2018 (6-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
1	precision	73.65	73.33	74.48	75.84	71.52	74.50	75.71	78.29	76.80	73.20	74.65	79.07	76.06	80.95	78.69
	recall	72.19	80.13	71.52	74.83	74.83	73.51	70.20	66.89	63.58	74.17	70.20	67.55	71.52	56.29	63.58
	f1 score	72.91	76.58	72.97	75.33	73.14	74.00	72.85	72.14	69.57	73.68	72.35	72.86	73.72	66.41	70.33
2	precision	83.05	80.77	78.87	80.45	79.53	79.43	79.66	78.69	72.37	77.94	78.17	68.94	67.60	74.29	83.19
	recall	64.90	69.54	74.17	70.86	66.89	74.17	62.25	63.58	72.85	70.20	73.51	73.51	80.13	68.87	65.56
	f1 score	72.86	74.73	76.45	75.35	72.66	76.71	69.89	70.33	72.61	73.87	75.77	71.15	73.33	71.48	73.33
4	precision	79.39	79.67	77.08	74.64	78.17	78.99	75.54	70.55	73.10	79.85	82.03	81.54	72.79	83.74	76.67
	recall	68.87	64.90	73.51	68.21	73.51	72.19	69.54	76.16	70.20	70.86	69.54	70.20	70.86	68.21	76.16
	f1 score	73.76	71.53	75.25	71.28	75.77	75.43	72.41	73.25	71.62	75.09	75.27	75.44	71.81	75.18	76.41
7	precision	77.30	76.77	73.79	75.52	81.68	74.50	75.71	78.29	76.80	73.20	76.81	77.44	72.03	79.84	76.92
	recall	71.19	78.81	70.86	71.52	70.86	73.51	70.20	66.89	63.58	74.17	70.20	68.21	68.21	65.56	66.23
	f1 score	74.66	77.78	72.30	73.47	75.89	74.00	72.85	72.14	69.57	73.68	73.36	72.54	70.07	72.00	71.17

Figure 4. Performances of Selected Best Models for the Usage Class

Similar		Jiang2022 (11-class)					Jurgens2018 (7-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
2	precision	57.45	46.94	65.79	60.00	67.50	62.07	53.12	54.29	59.65	63.46
	recall	64.29	54.76	59.52	71.43	64.29	58.06	54.84	61.29	54.84	53.23
	f1 score	60.67	50.55	62.50	65.22	65.85	60.00	53.97	57.58	57.14	57.89
5	precision	65.12	54.90	63.89	61.36	58.00	63.46	63.27	69.77	57.14	57.63
	recall	66.67	66.67	54.76	64.29	69.05	53.23	50.00	48.39	51.61	54.84
	f1 score	65.88	60.22	53.97	62.79	63.04	57.89	55.86	57.14	54.24	56.20
6	precision	58.14	56.82	58.54	62.79	60.53	59.32	70.00	60.71	66.67	81.08
	recall	59.52	59.52	57.14	64.29	54.76	56.45	45.16	54.84	58.06	48.39
	f1 score	58.82	58.14	57.83	63.53	57.50	57.85	54.90	57.63	62.07	60.61
11	precision	50.91	53.19	61.22	59.52	56.82	64.29	50.75	55.93	57.14	56.67
	recall	66.67	59.52	71.43	59.52	59.52	58.06	54.84	53.23	45.16	54.84
	f1 score	57.73	56.18	65.93	59.52	58.14	61.02	52.71	54.55	50.45	55.74

Figure 5. Performances of Selected Best Models for the Similarity Class

Background		Jiang2022 (11-class)					Jiang2022 (7-class)					Jurgens2018 (6-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
2	precision	75.50	76.14	76.04	78.85	75.43	82.17	81.27	78.02	80.72	78.59	78.06	76.74	75.08	74.34	74.34
	recall	77.82	68.60	74.74	69.97	75.43	82.87	79.21	81.74	82.30	78.37	76.85	78.40	74.38	77.78	77.78
	f1 score	76.64	72.17	75.39	74.14	75.43	82.52	80.23	79.84	81.50	78.48	77.45	77.56	74.73	76.02	76.02
6	precision	76.36	71.78	72.78	78.75	75.77	82.07	82.91	80.11	79.94	76.07	78.22	76.38	76.88	73.08	76.56
	recall	71.67	79.86	78.50	77.13	75.77	75.84	81.74	83.71	80.62	84.83	78.70	76.85	75.93	82.10	75.62
	f1 score	73.94	75.61	75.53	77.93	75.77	78.83	82.32	81.87	80.28	80.21	78.46	76.62	76.40	77.33	76.09
8	precision	75.60	73.61	75.95	78.07	69.64	79.53	76.49	78.17	80.50	76.62	76.05	76.98	72.30	76.88	71.43
	recall	75.09	72.35	75.43	71.67	79.86	86.24	86.80	81.46	81.18	82.87	72.53	69.14	76.54	75.93	78.70
	f1 score	75.34	72.98	75.68	74.73	74.40	82.75	81.32	79.78	80.84	79.62	74.25	72.85	74.36	76.40	74.89
11	precision	74.38	73.11	76.22	73.75	75.68	77.15	78.98	82.20	77.40	81.65	67.94	76.20	75.00	76.19	75.37
	recall	82.35	76.11	79.86	75.77	76.45	88.20	78.09	77.81	83.71	75.00	82.41	78.09	80.56	74.07	79.32
	f1 score	78.12	74.58	78.00	74.75	76.06	82.31	78.53	79.94	80.43	78.18	74.48	77.13	77.68	75.12	77.29

Figure 6. Performances of Selected Best Models for the Background Class

Future		Jiang2022 (11-class)					Jiang2022 (7-class)					Jurgens2018 (6-class)				
ID	metric	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353	5171	13249	25603	32491	47353
2	precision	87.50	84.62	66.67	76.47	72.22	75.00	70.59	92.31	72.22	76.47	75.00	100.00	87.50	81.25	76.47
	recall	82.35	64.71	94.12	76.47	76.47	70.59	70.59	70.59	76.47	76.47	70.59	64.71	82.35	76.47	76.47
	f1 score	84.85	73.33	78.05	76.47	74.29	72.73	70.59	80.00	74.29	76.47	72.73	78.57	84.85	78.79	76.47
5	precision	82.35	88.24	68.42	73.68	81.25	86.67	83.33	72.22	92.86	86.67	92.31	72.22	81.25	85.71	100.00
	recall	82.35	88.24	76.47	82.35	76.47	76.47	58.82	76.47	76.47	76.47	70.59	76.47	76.47	70.59	82.35
	f1 score	82.35	88.24	72.22	77.78	78.79	81.25	68.97	74.29	83.87	81.25	80.00	74.29	78.79	77.42	90.32
7	precision	76.47	65.00	92.86	72.22	85.71	76.47	80.00	68.75	77.78	73.33	68.42	68.42	68.42	86.67	100.00
	recall	76.47	76.47	76.47	76.47	70.59	76.47	70.59	64.71	82.35	64.71	76.47	76.47	76.47	76.47	76.47
	f1 score	76.47	70.27	83.87	74.29	77.42	76.47	75.00	66.67	80.00	68.75	72.22	72.22	72.22	81.25	86.67
8	precision	92.86	81.25	80.00	93.75	92.86	80.00	66.67	76.47	66.67	82.35	86.67	81.25	76.47	65.00	81.25
	recall	76.47	76.47	70.59	88.24	76.47	70.59	82.35	76.47	82.35	82.35	76.47	76.47	76.47	76.47	76.47
	f1 score	83.87	78.79	75.00	90.91	83.87	75.00	73.68	76.47	73.68	82.35	81.25	78.79	76.47	70.27	78.79

Figure 7. Performances of Selected Best Models for the Future Work Class

3.2. Semantic Main Path Network Extraction

3.2.1. Citation Network Building

We built a series of semantic citation networks based on citation function classification results. Starting from an empty citation network, a citation edge was added between a pair of papers if there existed **at least one** in-text citation about extension **or** motivation (`add_Ext_Mot`) using the "best" *extension* or *motivation* models selected in the recall-oriented approach in Sect. 3.1. Taking the same recall-oriented approach, more citation edges were added if there existed at least one *usage* citation between a pair of papers (`add_Ext_Mot_Use`). Similarly, the semantic citation network was further expanded with citation edges based on the chosen *similarity* model (`add_Ext_Mot_Use_Sim`). On the other hand, we also built the fourth semantic citation network by deleting incidental in-text citations from the original citation network. For each pair of papers, if **all** in-text citations between them were *neutral* or *future work* citations, the citation edge was removed from the citation network (`del_Bkg_Fut`).

3.2.2. Main Path Network Extraction

The citation networks we analysed have many small strongly connected components (SCC) so we used Jiang et al.'s Simple Search Path Count (SimSPC) approach (Jiang et al., 2020) for main path network extraction. For each edge (u, v), denote $N^-(u,v)$ and $N^+(u,v)$ as the number of search paths from source nodes to v through u and the number of search paths from u through v to sink nodes respectively, calculated in equations below.

$$N^-(u,v) = \sum_{\forall x \text{ in the SCC of } u} \left[P(x \rightarrow_{SCC} u | v) \times \sum_{\forall p \text{ s.t. } (p,\ u) \text{ is an edge}} N^-(p,x) \right]$$

$$N^+(u,v) = \sum_{\forall y \text{ in the SCC of } v} \left[P(v \rightarrow_{SCC} y | u) \times \sum_{\forall q \text{ s.t. } (y,\ q) \text{ is an edge}} N^-(y,q) \right]$$

where $P(x \rightarrow_{SCC} u | v)$ denotes the number of search paths from x to u inside the SCC of u *excluding* v and $P(v \rightarrow_{SCC} y | u)$ denotes the number of search paths from v to y inside the SCC of v *excluding* u. Then SimSPC is calculated as below.

$$\text{SimSPC}(u,v) = \begin{cases} \sum_{P^{SCC}_{x \rightarrow (u,v) \rightarrow y}} \left[\sum_{(p,x) \text{ outside } SCC} N^-(p,x) \times \sum_{(y,q) \text{ outside } SCC} N^+(y,q) \right] & \text{if } (u,v) \text{ in same } SCC \\ N^-(u,v) \times N^+(u,v) & \text{otherwise} \end{cases}$$

where $P^{SCC}_{x \rightarrow (u,v) \rightarrow y}$ is the number of simple paths from x to y through edge (u, v) in their common SCC.

Based on SimSPC, Jiang et al. also proposed an approximate data structure to hold the intermediate status of main path exploration. We used their `JPMA` (Java package for Main Path Analysis) package for main path network extraction. Following Jiang et al. (2020), we split the papers of each area of analysis into several time segments, built citation networks for each time segment, extracted top-*K* key-route main paths (Liu et al., 2012) on each time segment, and merged them into the unified *main path network*.

4. Experimental Analysis

We present two cases studies to demonstrate the superiority of the semantic main path network approach. The citation networks were extracted from the 2015 version of ACL Anthology Network (AAN; Radev et al., 2013) about two computational linguistics areas, natural language parsing[3] (`AANPar`) and automatic document summarisation (`AANSum`). The experimental setup is detailed in Supplementary Sect. B. Following the common practice in MPA, semantic main path networks were extracted from the largest connected components. Key-route MPA was used for main path extraction.

4.1. Case Study 1: Natural Language Parsing

4.1.1. Main Path Network

For comparison purpose, Figure 8 presents the main path network extracted from the original citation network `AANPar`. Topic branches are numbered. Seminal papers (verified according to the authors' knowledge about the domain) are in red rectangles, while survey-style papers are in ovals, such as special issue or shared task introduction papers. Table 3 shows the representative main path papers on each topic branch (See Supplementary Table S5 for a more complete list), where topic keywords and short excerpts for certain papers are to assist understanding. Branch 1 is about the early studies focusing on various types of grammatical formalisms[4], such as **categorical grammar**, **unification grammar**, **categorical unification grammar** and **Lambek calculus** etc. However, since late 1980s, the domain started to have a sense of probabilistic thinking (Branch 2). Branch 3 shows the important development where **Penn TreeBank** (J93-2004 and H94-1020) appeared as the most important linguistic resource that most future papers used for developing and evaluating parsing techniques. Branch 4 represents the mainstream of statistical parsing in the 1990s and 2000s, such as **maximum entropy modelling** (W96-0213, A00-2018) or in another name **log-linear model** (P04-1014), **conditional random fields** (N03-1028), and **max-margin parsing** (W04-3201,

[3] Parsing: Parsing, syntax analysis, or syntactic analysis is the process of analyzing a string of symbols, either in natural language, computer languages or data structures, conforming to the rules of a formal grammar. See Wikipedia page: https://en.wikipedia.org/wiki/Parsing.
[4] Note that more grammars were proposed even earlier, outside our time range of analysis.

P05-1012). C96-1025 (by J. M. Eisner) and P97-1003 (by M. Collins) are also two important works. Note that, C00-1011 and P00-1009 were two papers on **Data-Oriented Parsing (DOP)** promoted by Rens Bod, which however ceased in the wave of statistical parsing dominated by other proposals presented above. Early studies about **dependency analysis** blossomed into the huge Branch 5 and became the dominant trend since around 2005, further expediated by two important shared tasks W06-2920 and D07-1096, which then diverted into Branch 6 about dependency parsing of **morphologically rich languages** and Branch 7 about **cross-lingual dependency parsing**.

An issue was that many main path papers were connected by incidental citations. For instance, the citation between Rens Bod's DOP paper C00-1011 and Eugene Charniak's seminal paper A00-2018 ("A Maximum-Entropy-Inspired Parser") said that C00-1011 "stays behind the scores of" A00-2018, a weak citation about performance comparison. For another instance, H91-1037 received only 10 citations in the `AAN` dataset (1985-2015). It was included only because the citing journal paper J93-2004 received 1006 citations, resulting in a very high SPC(H91-1037, J93-2004), but the citation was incidental, shown by the citation context excerpt below (cited paper underlined).

"5.2 Future Directions A large number of research efforts, …, have relied on the output of the Penn Treebank Project to date. A few examples already in print: a number of projects investigating stochastic parsing have used either the POS-tagged materials (…) or the skeletally parsed corpus (<u>Weischedel et al. 1991</u>; …)."

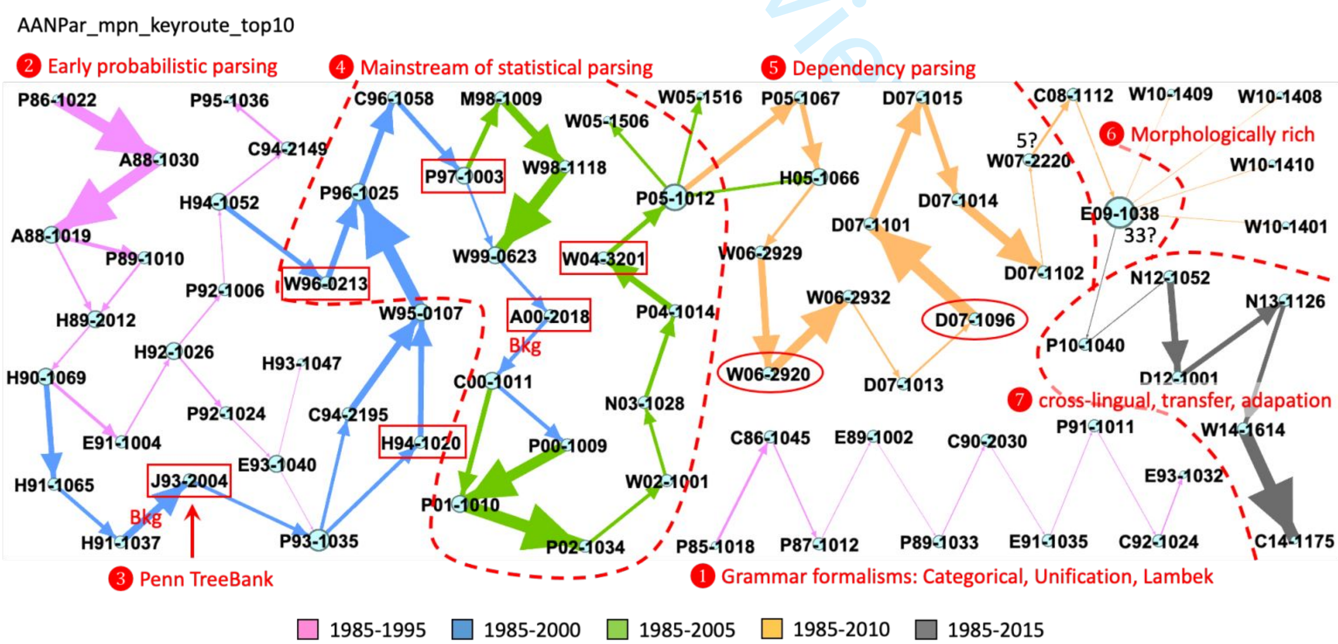

Figure 8. Main Path Network of the Parser Network `AANPar`

Table 3. Main Path Network Papers of the Parser Network `AANPar`

ACLID	**Title**
Branch 1	
P85-1018	Using Restriction To Extend Parsing Algorithms For **Complex-Feature-Based Formalisms**
C86-1045	**Categorial Unification Grammars**
P87-1012	A Lazy Way To Chart-Parse With **Categorial Grammars**
C90-2030	Normal Form Theorem Proving For The **Lambek Calculus**
P91-1011	Efficient Incremental Processing With **Categorial Grammar**
C92-1024	Chart Parsing **Lambek Grammars**: Modal Extensions And Incrementality
Branch 2	
A88-1030	Finding Clauses In Unrestricted Text By Finitary And **Stochastic** Methods
A88-1019	A **Stochastic** Parts Program And Noun Phrase **Parser** For Unrestricted Text
H92-1026	Towards History-Based Grammars: Using Richer Models For **Probabilistic Parsing**
P92-1024	Development And Evaluation Of A Broad-Coverage **Probabilistic Grammar** Of English-Language Computer Manuals
E93-1040	Parsing The Wall Street Journal With The Inside-Outside Algorithm
	Excerpt: We report grammar inference experiments on partially parsed sentences taken from the **Wall Street Journal corpus** using the inside-outside algorithm for **stochastic context-free grammars**.
Branch 3	
J93-2004	**Building A Large Annotated Corpus Of English: The Penn Treebank**
P93-1035	Automatic Grammar Induction And Parsing Free Text: A **Transformation-Based** Approach
	Excerpt: All of the experiments presented below were done using the **Penn Treeban**k annotated corpus (MSM93). (MSM93: J93-2004)
H94-1020	**The Penn Treebank: Annotating Predicate Argument Structure**
W95-1017	Text Chunking Using **Transformation-Based** Learning
Branch 4	
W96-0213	A **Maximum Entropy** Model For Part-Of-Speech Tagging
C96-1058	Three New **Probabilistic Models** For *Dependency Parsing*: An Exploration
P97-1003	**Three Generative Lexicalized Models For Statistical Parsing**
A00-2018	A **Maximum-Entropy-Inspired** Parser
C00-1011	Parsing With The Shortest Derivation (about **DOP by** Rens Bod)
P00-1009	An Improved Parser For **Data-Oriented** Lexical-Functional Analysis (about **DOP by** Rens Bod)
N03-1028	Shallow Parsing With **Conditional Random Fields**
P04-1014	Parsing The WSJ Using **CCG** And **Log-Linear Models**
W04-3201	**Max-Margin Parsing**
P05-1012	**Online Large-Margin Training** Of *Dependency Parsers*
Branch 5	
H05-1066	Non-Projective **Dependency Parsing** Using Spanning Tree Algorithms
W06-2920	**CoNLL-X Shared Task On Multilingual Dependency Parsing**
W06-2932	**Multilingual Dependency** Analysis With A Two-Stage Discriminative Parser
D07-1096	**The CoNLL 2007 Shared Task on Dependency Parsing**
D07-1101	Experiments with a Higher-Order Projective **Dependency Parser**
D07-1014	Probabilistic Models of **Nonprojective Dependency Trees**
D07-1102	**Log-Linear** Models of **Non-Projective Trees k-best MST Parsing** and Tree-Ranking
Branch 6	
W10-1401	Statistical Parsing of **Morphologically Rich** Languages (SPMRL) What How and Whither
W10-1410	Lemmatization and Lexicalized Statistical Parsing of **Morphologically-Rich** Languages: the Case of **French**
Branch 7	
N12-1052	**Cross-lingual** Word Clusters for Direct **Transfer** of Linguistic Structure
D12-1001	**Syntactic Transfer** Using a **Bilingual** Lexicon
N13-1126	**Target Language Adaptation** of Discriminative **Transfer** Parsers
W14-1614	Treebank Translation for **Cross-Lingual** Parser Induction
C14-1175	Rediscovering Annotation Projection for **Cross-Lingual** Parser Induction

4.1.2. Semantic Main Path Network: Add Extension and Motivation Citations

The above observations motived us to exploit the semantic relationships between papers in MPA. Figure 9-12 show the main path networks extracted from the four parser networks filtered by adding or deleting in-text citations of certain citation functions (Sect. 3.1), named `AANPar_add_Ext_Mot`, `AANPar_add_Ext_Mot_Use`, `AANPar_add_Ext_Mot_Use_Sim`, and `AANPar_del_Bkg_Fut`. We were excited to see wonderful chemical reactions when main path analysis met citation function classification. Each semantic main path network revealed some novel branches or new papers. They collectively drew a more comprehensive picture of domain development.

The `AANPar_add_Ext_Mot` network (Figure 9, Table 4 and Supplementary Table S6 for a more complete list) detected the early development of parsing technology, e.g. Branch 2 about old parsers such as **shift-reduce parsing**, **left-corner parsing**, **tabular parsing**, and **left-to-right (LR) parsing** etc (a new branch). Similarly, we could also see an (isolated) early development of **probabilistic** approaches (Branch 3; details in Supplementary Table S5). Another knowledge source of statistical parsing started from E85-1024 (“A probabilistic parser”) to J94-2001 (“Tagging English Text with a Probabilistic Model”) and W96-0213, then through P02-1034 into Branch 4 about **multiple parse ranking and re-ranking**, a new branch detected. A00-2018 (“A Maximum-Entropy-Inspired Parser”) was the third source of the statistical parsing mainstream. The fourth source (Branch 5) started form C92-2065 and C92-2066 about Probabilistic/Stochastic Tree-Adjoining Grammar to W00-1201 (on Chinese TreeBank). From the right part of Figure 8, we could see a branch of **DOP** papers published by Rens Bod until P01-1010. It was gradually merged into the dominant Branch 7 about dependency parsing, following a similar evolution pathway as in Figure 7 into the shared task on dependency parsing (D07-1096). D08-1059 (“A Tale of Two Parsers: Investigating and Combining Graph-based and Transition-based Dependency Parsing”) was motivated (denoted by “Mot” on the edge) by two papers P07-1050 (“K-best Spanning Tree Parsing”) and D07-1013 (“Characterizing the Errors of Data-Driven Dependency Parsing Models”). See Supplementary Sect. D for the citation context excerpts.

Note that, there was an interesting but potentially problematic Branch 8 about machine translation (MT) using dependency parsing. Though looking topically relevant, it is unclear whether H05-1066 indeed extended P05-1012 (denoted by “Ext” on the edge). The citation context excerpt below reveals that although “improving upon” may indicate an extension, the whole context may also be recognised as “Similar” or even “CoCoGM”. This example implies that multi-label citation function classification might be a promising future direction to explore (Lauscher at al., 2021).

"We mentioned above that our approach appears to be similar to that of reranking for statistical parsing (Collins, 2000; Charniak and Johnson, 2005). While it is true that we are improving upon the output of the automatic parser, we are not considering multiple alternate parses."

There are of course some vague cases, such as the citation from C02-1126 to W00-1201, a self-citation by D. M. Bikel and D. Chiang. From the citation context below, our selected model might have selected expressions like "starting from" and "we have modified" as strong signals for extension class ("Ext"). Supplementary Sect. D presents more citation context excerpts to help readers understand the citation functions marked on certain edges.

"The third experiment was on the Chinese Treebank, starting with the same head rules used in (Bikel and Chiang, 2000). These rules were originally ..., and although we have modified them for parsing, ..."

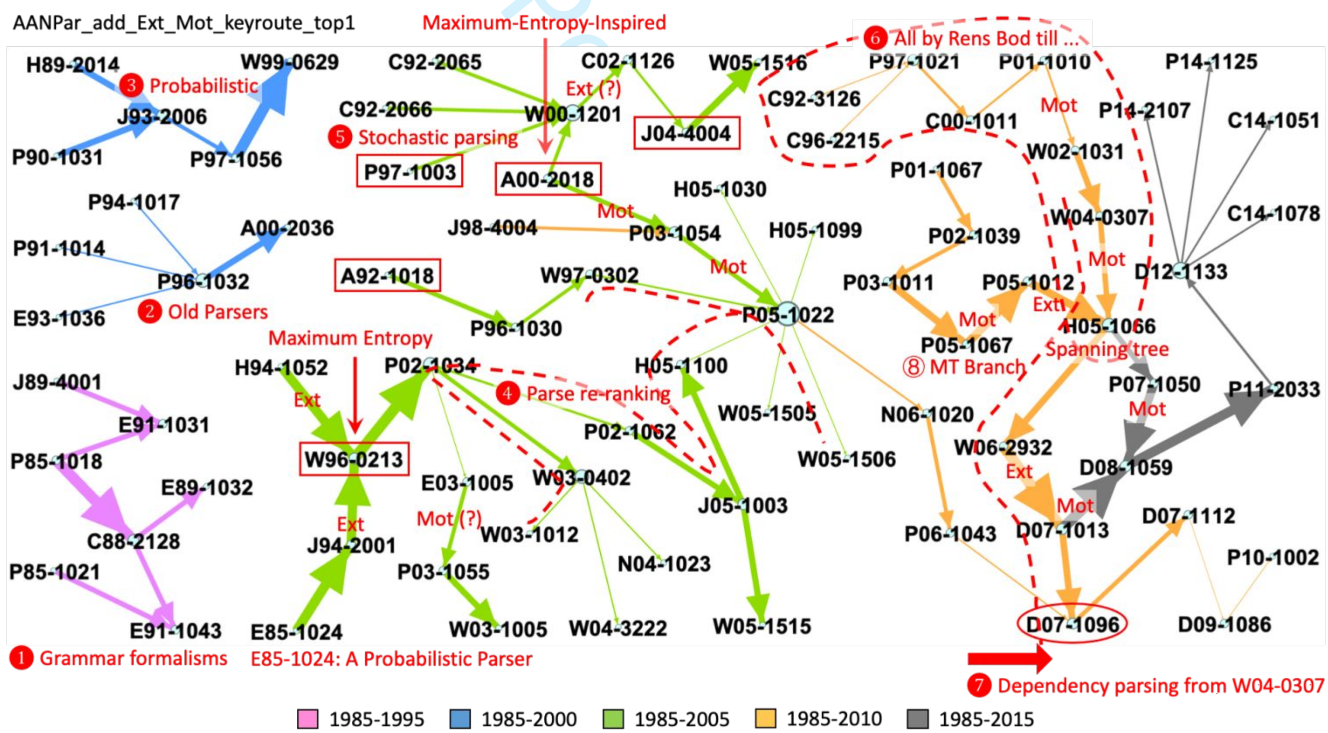

Figure 9. Main Path Network of the Parser Network `AANPar` by Adding Extension and Motivation Citations

Table 4. Main Path Network Papers of the Parser Network `AANPar_Add_Ext_Mot`

ACLID	Title
Branch 2	
P91-1014	Polynomial Time And Space **Shift-Reduce Parsing** Of Arbitrary **Context-Free Grammars**
E93-1036	Generalized **Left-Corner Parsing**
P94-1017	An Optimal **Tabular Parsing** Algorithm
P96-1032	Efficient **Tabular L R Parsing**
A00-2036	**Left-To-Right Parsing** And Bilexical **Context-Free Grammars**
Branch 4	
W97-0302	Global Thresholding And **Multiple-Pass Parsing**
P02-1034	New **Ranking Algorithms** For Parsing And Tagging: Kernels Over Discrete Structures And The Voted Perceptron
P02-1062	**Ranking Algorithms** For Named Entity Extraction: Boosting And The Voted Perceptron
J05-1003	**Discriminative Reranking** For Natural Language Parsing
P05-1022	Coarse-To-Fine **N-Best Parsing** And MaxEnt **Discriminative Reranking**
…	
Branch 5	
C92-2065	**Probabilistic Tree-Adjoining Grammar** As A Framework For **Statistical** Natural Language Processing
C92-2066	**Stochastic Lexicalized Tree-Adjoining Grammars**
…	
P97-1003	Three Generative, Lexicalised Models for **Statistical Parsing** (by Michael Collins)
W00-1201	Two Statistical Parsing Models Applied To The ***Chinese Treebank***
J04-4004	Intricacies Of **Collins Parsing** Model Talks about P97-1003
Branch 6	(By Rens Bod on Data-Oriented Parsing, DOP)
C92-3126	A Computational Model Of Language Performance: ***Data Oriented Parsing***
P97-1021	A ***DOP*** Model For Semantic Interpretation
C00-1011	Parsing With The Shortest Derivation Exceprt: To investigate this question we created a new **STSG-DOP** model which uses this bias as a feature. This non-probabilistic **DOP** model parses each sentence by …
P01-1010	What Is The Minimal Set Of Fragments That Achieves Maximal Parse Accuracy?
Branch 8	(A “wrong” branch)
P01-1067	A *Syntax-Based Statistical Translation Model*
P02-1039	A Decoder For *Syntax-Based Statistical MT*
P03-1011	Loosely *Tree-Based* Alignment For *Machine Translation*
P05-1067	*Machine Translation* Using Probabilistic Synchronous **Dependency** Insertion Grammars

4.1.3. Semantic Main Path Network: Further Add Usage and Similarity Citations

By further adding usage citations, i.e. on `AANPar_add_Ext_Mot_Use`, we could see drastically richer diversity in the development branches (Figure 10, Table 5 and Supplementary Table S7). Again, statistical parsing techniques evolved from multiple intelligent sources (Branches 1-3). Now we could see a clear notion of "**corpus-based**" parsing (Branch 1). Branch 2 was motivated by H93-1047 ("Automatic Grammar Induction And Parsing Free Text: A Transformation-Based Approach", seemingly the same paper as P93-1035) and developed into "**shallow parsing**" of words into "**text chunks**"[5]. Similar to Figure 9, Branch 3 started from the early probabilistic parser (E85-1024) and developed through W96-0213 to J04-4004, but this time the seminal paper J93-2004 about the **Penn Treebank** project emerged. Most subsequent papers used Peen Treebank in developing and evaluating parsing methods. Another knowledge source originated from the **corpus-based** statistical parsing branch, through J95-4004 and J93-2006, and joined Branch 3 at P01-1067, the paper about syntax-based statistical translation as the start of the MT branch. The citation context excerpt below shows that P01-1067 used J95-4004.

> *"Brill's part-of-speech (POS) tagger (Brill, 1995) and Collins' parser (Collins, 1999) were used to obtain parse trees for the English side of the corpus."*

Considering the DOP branch lead by Rens Bod from J93-2006 into C00-1011 (refer to Branch 6 in Table 4), it "developed" into Branch 4 and found an important shared task W05-0620 on **semantic role labelling** (SRL; prediction of predicate arguments). Then this branch "vanished". This was understandable because SRL became a rather standalone area since then (just like dependency parsing)[6], so it was possible that the SRL papers became less related to, i.e. citing less or less cited by, parsing papers. The dependency parsing branch also evolved into a similar Branch 6 about **cross-lingual dependency parsing** as in Figure 8-9 but embraced a more diverse set of papers.

By further adding similarity citations, i.e. on the `AANMPar_add_Ext_Mot_Use_Sim` network, the main path network extracted was quite similar (Figure 11, Table 6 and Supplementary Table S8). However, we could observe quite a few interesting new branches emerged. Starting from the seminal Penn Treebank paper J93-2004, two new branches developed from P97-1062 and W97-0301 based on usage citations respectively into Branch 1 about **rhetorical parsing** and Branch 2 about **probabilistic parsing with CCG (Combinatory Categorial Grammar)**. Through similarity citations, we found some

---

[5] From Wikipedia: Shallow parsing (also chunking or light parsing) is an analysis of a sentence which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings (noun groups or phrases, verb groups, etc.).
[6] Both semantic role labelling and dependency parsing became rather standalone topics and had bespoke monographs on these two topics.

new main path papers, such as J96-1002 ("A Maximum Entropy Approach to Natural Language Processing") which was heavily cited (387 times).

"The maximum entropy models used here are similar in form to those in (Ratnaparkhi, 1996; Berger, Della Pietra, and Della Pietra, 1996; Lau, Rosenfeld, and Roukos, 1993)."

We also found a small new branch from J95-4004 to P98-1034 based on similarity citation. The following citation context excerpt proved that similarity citation is indeed relevant to the knowledge flow of scientific ideas.

"Our benefit measure is identical to that used in transformation-based learning to select an ordered set of useful transformations (Brill, 1995)."

The domain then evolved to the dominant dependency parsing branch (Branch 3). Besides the two important shared tasks W06-2920 and D07-1096, we were excited to see two new shared tasks W08-2121 and W09-1201 about **joint syntactic and semantic dependency parsing**, and the subsequent studies on **semantic dependency parsing** (W09-1208 and D09-1004).

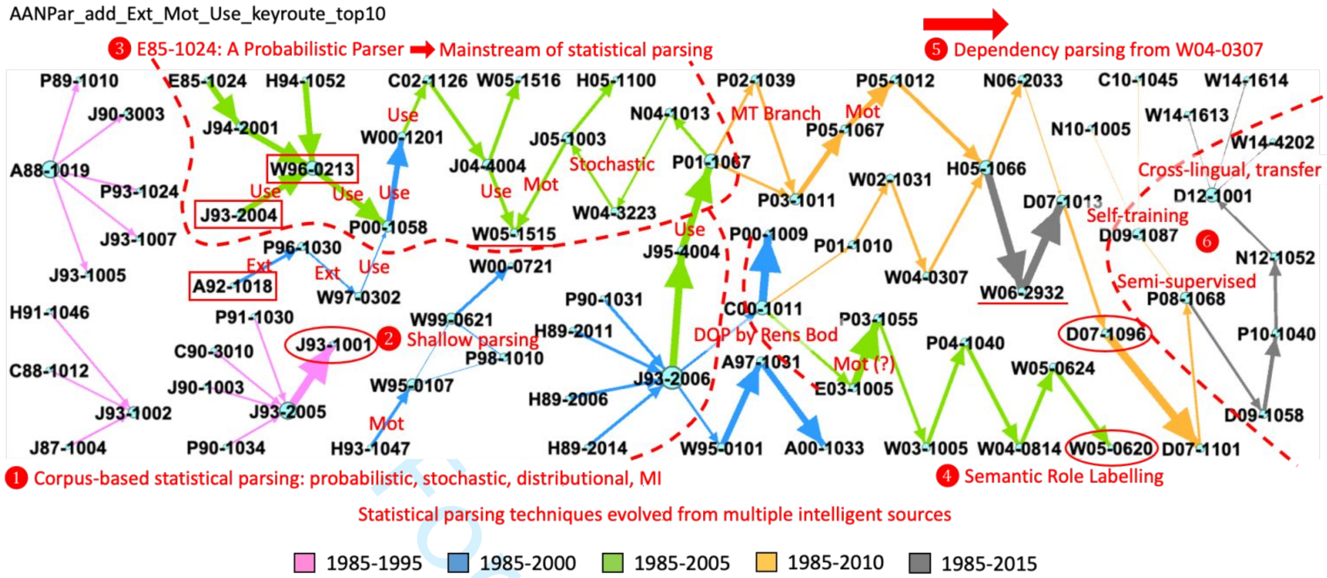

Figure 10. Main Path Network of the Parser Network `AANPar` by Adding Extension, Motivation and Usage Citations

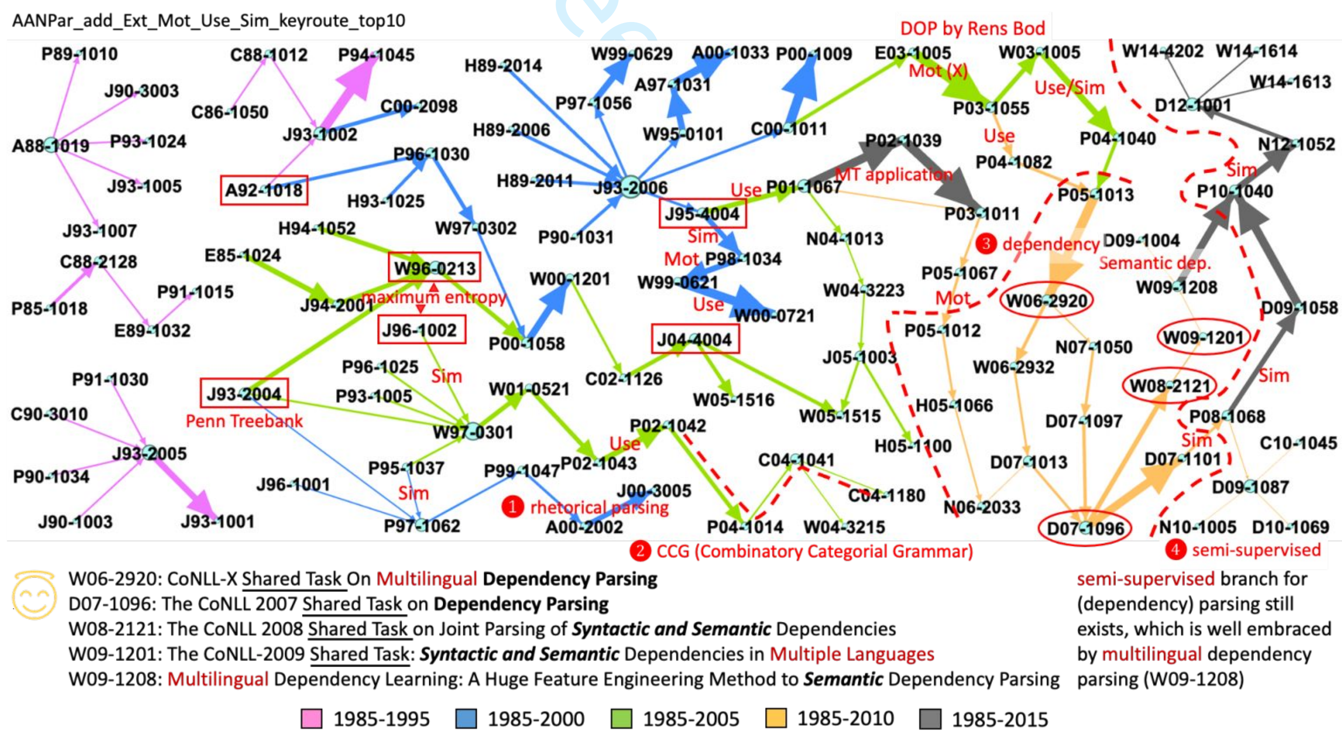

Figure 11. Main Path Network of the Parser Network `AANPar` by Adding Extension, Motivation, Usage and Similarity Citations

Table 5. Main Path Network Papers of the Parser Network `AANPar_Add_Ext_Mot_Use`

ACLID	Title
Branch 1	
J93-1002	Generalized **Probabilistic LR Parsing** Of Natural Language (**Corpora**) With Unification-Based Grammars
J93-1001	**Introduction To The Special Issue On Computational Linguistics Using Large Corpora**
P90-1031	Parsing The LOB **Corpus**
Branch 2	
W95-0107	**Text Chunking** using Transformation-Based Learning
W99-0621	A Learning Approach to **Shallow Parsing**
W00-0721	**Shallow Parsing** by Inferencing with Classifiers
Branch 3	
…	
J93-2004	**Building A Large Annotated Corpus Of English: The Penn <u>Treebank</u>**
…	
W00-1201	Two **Statistical Parsing** Models Applied To The ***Chinese Treebank***
…	
P01-1067	A Syntax-Based *Statistical Translation Model*
Branch 4	
W04-0814	The University Of Amsterdam At Senseval-3: **Semantic Roles** And Logic Forms
W05-0620	**Introduction To The CoNLL-2005 Shared Task: Semantic Role Labeling**
Branch 6	(Extended branch about semi-supervised learning and cross-lingual dependency parsing)
P08-1068	Simple **Semi-supervised** Dependency Parsing
D09-1087	**Self-Training** PCFG Grammars with Latent Annotations Across Languages
W14-1613	Distributed Word Representation Learning for **Cross-Lingual** Dependency Parsing
W14-4202	Learning from a Neighbor: Adapting a Japanese Parser for Korean Through Feature **Transfer Learning**

Table 6. Main Path Network Papers of the Parser Network `AANPar_Add_Ext_Mot_Use_Sim`

ACLID	Title
Branch 1	
P99-1047	A Decision-Based Approach To **Rhetorical Parsing**
J00-3005	The **Rhetorical Parsing** Of Unrestricted Texts: A Surface-Based Approach
Branch 2	
P02-1042	Generative Models For **Statistical Parsing** With **Combinatory Categorial Grammar**
P04-1014	Parsing The WSJ Using **CCG** And **Log-Linear Models**
C04-1180	Wide-Coverage Semantic Representations From A **CCG Parser**
Branch 3	(extended branch of dependency parsing)
…	
W06-2920	**CoNLL-X Shared Task On Multilingual Dependency Parsing**
D07-1096	**The CoNLL 2007 Shared Task on Dependency Parsing**
W08-2121	**The CoNLL 2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies**
W09-1201	**The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages**
Branch 4	(extended to semantic dependency parsing)
W09-1208	**Multilingual Dependency Learning**: A Huge Feature Engineering Method to **Semantic Dependency Parsing**
D09-1004	**Semantic Dependency Parsing** of **NomBank** and **PropBank**: An Efficient Integrated Approach via a Large-scale Feature Selection

4.1.4. Semantic Main Path Network: Delete Neutral and Future Work Citations

Finally, although the semantic main path network extracted from `AANPar_Del_Bkg_Fut` (Figure 12, Table 7 and Table S9 for a more complete list) demonstrated high similairty to Figure 8, we could observe some interesting branches or papers. Since P08-1068, the domain diverted into a new branch about **optimization techniques** used in parsing algorithms, such as **dynamic programming**, **integer linear programming** and **dual decomposition** (Branch 2). Branch 3 was a similar cross-lingual dependency parsing branch as in Figure 8-11. However, it later evolved into Branch 4 about parsing morphologically rich languages through a new **shared task** (W13-4917), thus provided a complementary view to Branch 6 in Figure 8. From the fact that dependency parsing was directed by important shared tasks, we postulate that it was meaningful to recognize this shared task as main path paper (thus more meaningful than Figure 8). Note that, the semantic coherence of the semantic main path network by deleting neutral and future work citations might be weaker than those extracted by adding semantically more significant citations. For example, N07-1069 only made a result comparison with W06-2928, therefore it is less confident to say scientific ideas flew through this path.

> *"Here we can compare directly with the best systems for this dataset in CoNLL-X. The best system (Corston-Oliver & Aue, 2006), …."*

In summary, we found that different semantic main path networks may introduce more semantically coherent topics. We also suggest that different types of networks be built and merged into a more comprehensive representation of scientific domain's topic evolution.

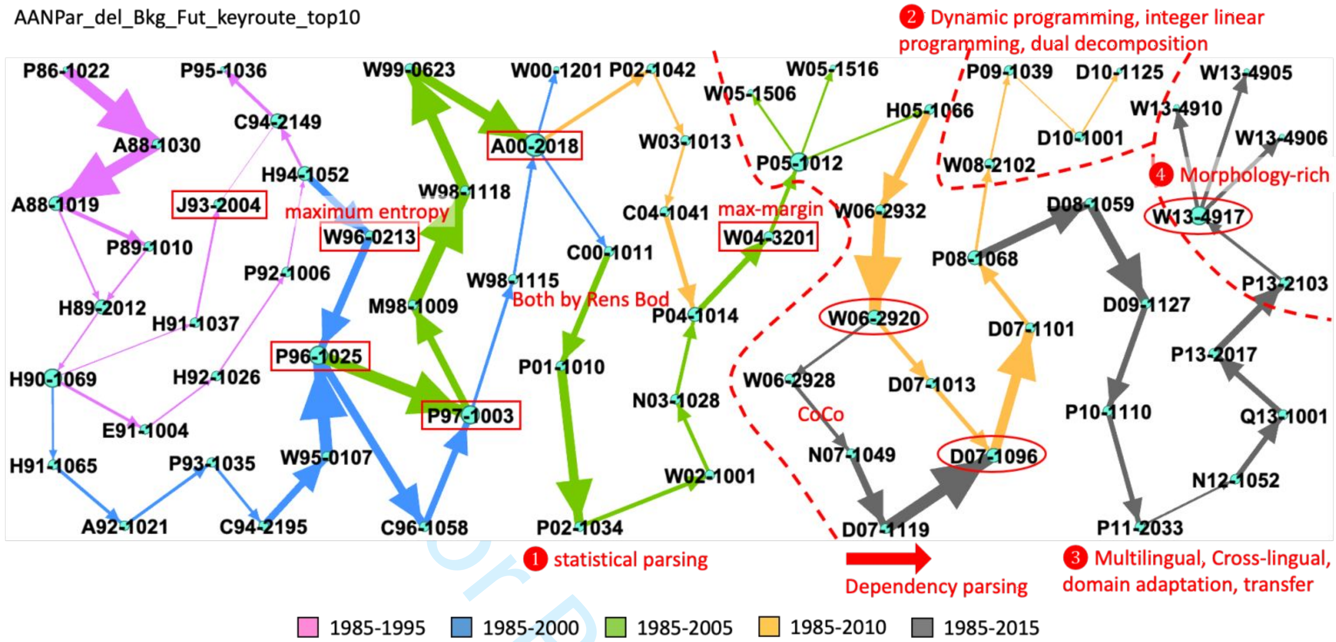

Figure 12. Main Path Network of the Parser Network `AANPar` by Deleting Background and Future Work Citations

Table 7. Main Path Network Papers of the Parser Network `AANPar_Del_Bkg_Fut` (~140)

ACLID	**Title**
Branch 2	
W08-2102	TAG, **Dynamic Programming**, and the Perceptron for Efficient, Feature-Rich Parsing
P09-1039	Concise **Integer Linear Programming** Formulations for Dependency Parsing
D10-1001	On **Dual Decomposition** and **Linear Programming** Relaxations for Natural Language Processing
D10-1125	**Dual Decomposition** for Parsing with Non-Projective Head Automata
Branch 3	
W06-2928	Dependency Parsing With Reference To Slovene **Spanish** And **Swedish**
N07-1049	Tree Revision Learning for Dependency Parsing
D07-1119	**Multilingual** Dependency Parsing and **Domain Adaptation** using DeSR
D09-1127	**Bilingually-Constrained** (Monolingual) Shift-Reduce Parsing
Q13-1001	Token and Type Constraints for **Cross-Lingual** Part-of-Speech Tagging
P13-2017	Universal Dependency Annotation for **Multilingual** Parsing
Branch 4	
P13-2103	A ***Unified Morpho-Syntactic*** Scheme of Stanford Dependencies
W13-4917	**Overview of the SPMRL 2013 Shared Task: A Cross-Framework Evaluation of Parsing Morphologically Rich Languages** W13-4905, W13-4906 and W13-4910 are all SPMRL 2013 Shared Task papers

4.2. Case Study 2: Automatic Document Summarisation

Due to space limit, an informative summary is presented here (Figure 13-17, Table 8). See Supplementary Sect. D for details of the main path papers. The main path network extracted from `AANSum` (Figure 12) covered a few early summarization studies centering around the usage of **semantic coherence** devices (Branch 1), such as **discourse structure**, **rhetorical relations**, and **lexical chains** (W97-0703: Using Lexical Chains For Text Summarization) etc. Then the main body of literature focused on **multidocument summarization** (Branch 2) pioneered by the seminal journal article J98-3005 ("Generating Natural Language Summaries From Multiple On-Line Sources"). The subsequent studies in this topic eventually gave birth to an important **Special Issue** on Summarization (J02-4001). Since the advent of PageRank in 1998, the **graph-based ranking** idea was introduced to the summarization domain for sentence ranking for extractive summarisation (Branch 3). Seminal works included P04-3020 ("Graph-Based Ranking Algorithms For Sentence Extraction Applied To Text Summarization"), W04-3252 ("TextRank: Bringing Order Into Texts"), the subsequent demonstration paper of TextRank (P05-3013), and its extension to multidocument summarization (I05-2004). More recently, a large body of the literature were about some interleaved topics: **optimization techniques** such as **submodular optimization** (E12-1023), **integer linear programming** (D12-1022), and **dual decomposition** (P13-1020); **compressive summarization** (P10-1058, P11-1049, D13-1047); and **compressive summarization based on dependency tree** (P14-2052, D14-196). Notably, comparison (sometimes weakness) function was the dominating citation function in Branch 4 in Figure 12. In addition, the only papers about summarization evaluation was N03-1020 about ROUGE ("Automatic Evaluation Of Summaries Using N-Gram Co-Occurrence Statistics"). These two drawbacks motivated us to explore novel branches of summarization using semantic main path networks.

By adding extension and motivation citations (Figure 14), we could see a larger early branch about the usage of rhetorical structure and found a seminal application in **scientific summarization** (J02-4002), which was extended by subsequent studies in other areas, like W03-0505 ("Summarising Legal Texts: Sentential Tense And Argumentative Roles"), evidenced by the citation context excerpt below.

> *"Our methodology builds and extends the Teufel and Moens (Teufel and Moens, 2002) approach to automatic summarization."*

In addition to the common topics like multidocument summarization (Branch 2) and graph-based ranking algorithms (Branch 5), we were also excited to see Branch 3 about **automatic evaluation** and related studies. Heavily cited ones included N03-1020 and W04-1013 about the ROUGE package. We also saw more studies about **sentence reduction**, **compression and**

fusion for summarization. Both Branch 4-1 and 4-2 were pioneered by K. R. McKewon in A00-1043 ("Sentence Reduction For Automatic Text Summarization"), A00-2024 ("Cut and Paste Based Text Summarization"), and J05-3002 ("Sentence Fusion For Multidocument News Summarization").

By further adding usage citations (Figure 15), although we lost the graph-based ranking branch (despite that we got a new paper W04-3247 about LexPageRank), we could uncover more novel topics and branches. Branch 2 about automatic evaluation included more important papers such as N04-1019 about the **Pyramid method** ("Evaluating Content Selection In Summarization: The Pyramid Method"). A significant new branch was Branch 3 about **scientific summarization** at right bottom, starting from the seminal paper J02-4002 to **citation function classification** (W06-1613, N07-1040) and **citation-based summarization** (C08-1087, N09-1066, P10-1057, and C10-1101). By further adding similarity citations (Figure 16), we could see one obvious expansion of Branch 1 about evaluation, starting from **factoid analysis** (W04-3254) to summarization **evaluation without human models**, including D09-1032 ("Automatically Evaluating Content Selection in Summarization without Human Models") and C10-2022 ("Multilingual Summarization Evaluation without Human Models"), both written by famous researchers in this domain (A. Nenkova and H. Saggion respectively).

Finally, the main path network extracted from `AANSum_del_Bkg_Fut` (Figure 17) recovered the vanished or shrunk branches about **multidocument summarization** (Branch 1) and **graph-based ranking** (Branch 2), and at the same time introduced some new papers (Table 8). It was again proved that, by gradually adding more semantics, the set of semantic main path networks together were more expressive than the semantics-agnostic counterpart.

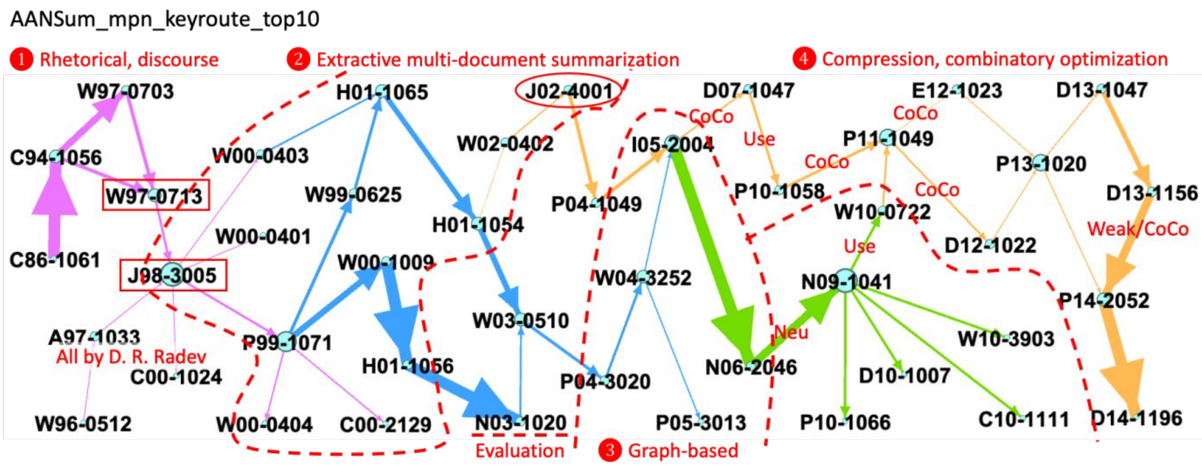

Figure 13. Main Path Network of the Summarisation Network `AANSum`

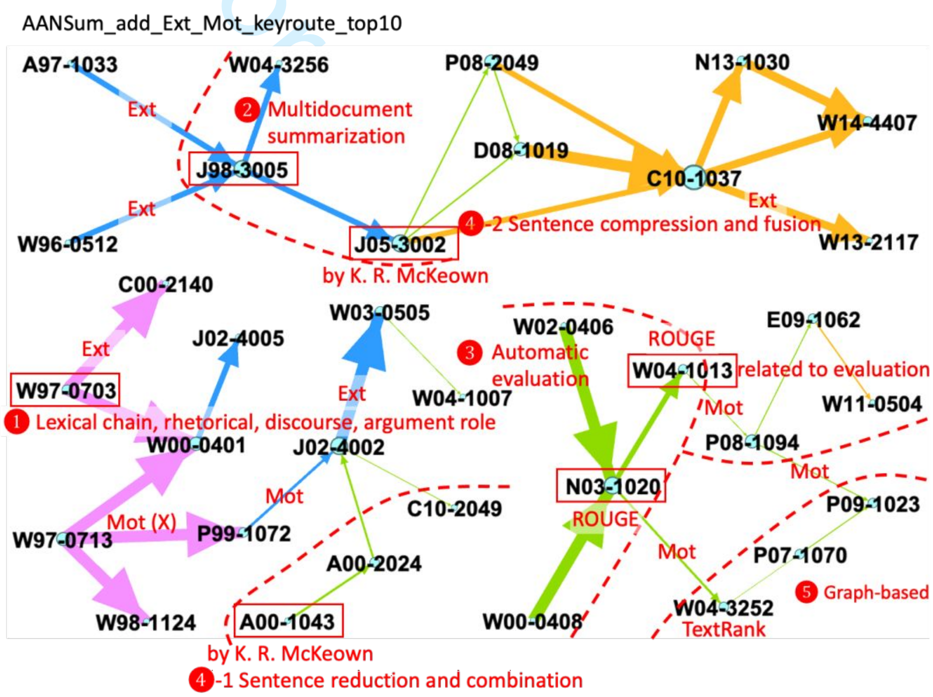

Figure 14. Main Path Network of the Summarisation Network `AANSum` by Adding Extension and Motivation Citations

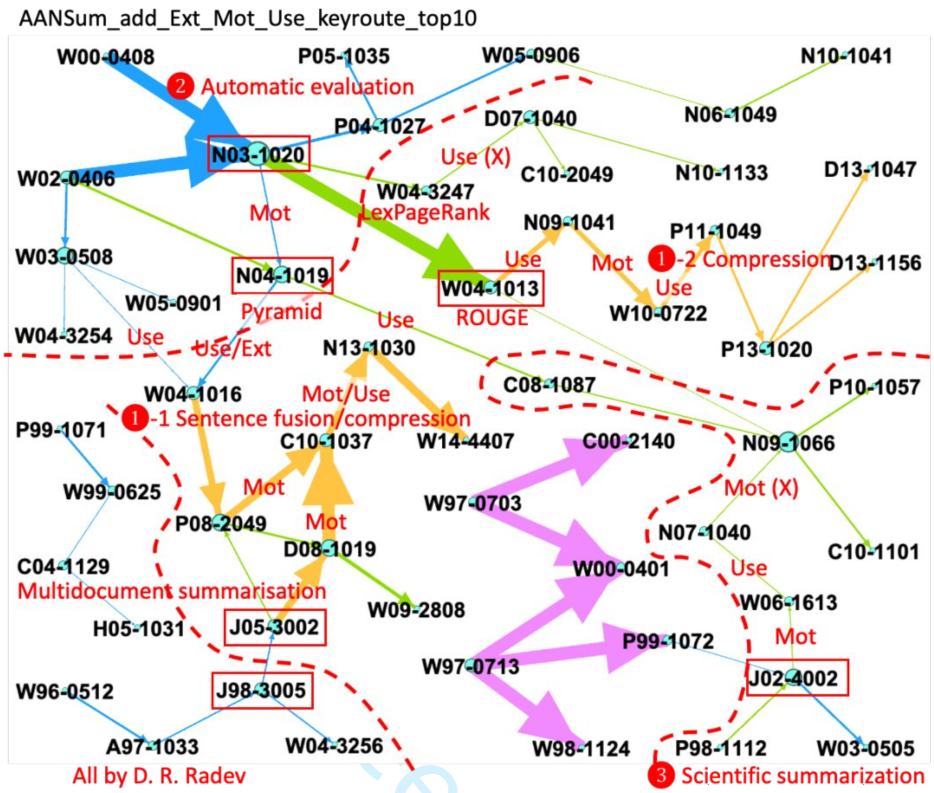

Figure 15. Main Path Network of the Summarisation Network `AANSum` by Adding Extension, Motivation and Usage Citations

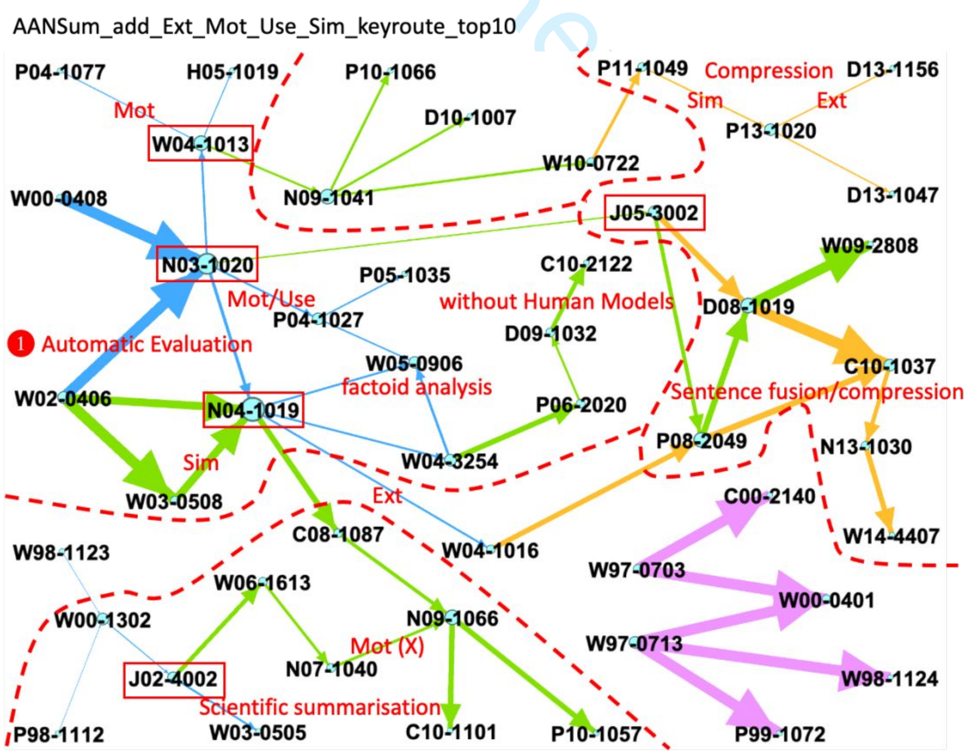

Figure 16. Main Path Network of the Summarisation Network `AANSum` by Adding Extension, Motivation, Usage and Similarity Citations

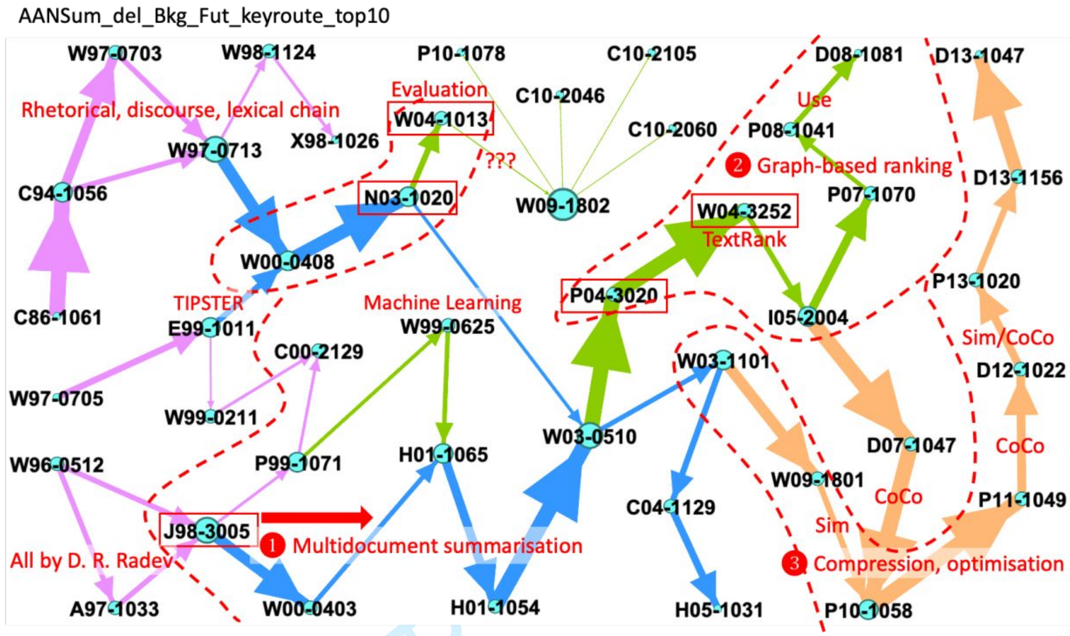

Figure 17. Main Path Network of the Summarisation Network `AANSum` by Deleting Background and Future Work Citations

Table 8. Representative Main Path Network Papers of the Summarisation Network `AANSum_Del_Bkg_Fut`

ACLID	Title
Branch 1	
W03-1101	Improving Summarization Performance By ***Sentence Compression*** - A Pilot Study
C04-1129	Syntactic ***Simplification*** For Improving Content Selection In **Multi-Document Summarization**
H05-1031	Automatically Learning Cognitive Status For **Multi-Document Summarization** Of Newswire
Branch 2	
P08-1041	Summarizing Emails with Conversational Cohesion and Subjectivity
	Abstract: Second, we use two **graph-based summarization approaches**, Generalized ClueWordSummarizer and PageRank, to extract sentences as summaries.
Branch 3	
W09-1802	A Scalable Global Model for Summarization
	We present an **Integer Linear Program** for exact inference under a maximum coverage model for automatic summarization.
C10-2105	Opinion Summarization with **Integer Linear Programming** Formulation for Sentence Extraction and Ordering
C10-2046	Learning to Model Domain-Specific Utterance Sequences for Extractive Summarization of Contact Center Dialogues
	Same authors as C10-2105: our method significantly outperforms competitive baselines based on the maximum coverage of important words using **integer linear programming**

5. Conclusions

This paper advocated a novel semantic main path network approach for extracting the scientific backbone from a citation network based on citation context analysis. First, the best models for extension, motivation, usage, similarity, neutral (equiv. background) and future work citation functions were picked, based on per-class performances analysis, from 55 contextualised citation function classification models trained from 11 model architectures based on SciBERT. Then, four semantic citation networks were created by gradually adding or deleting citations of certain functions. Citation addition was done on extension, motivation, usage and similarity functions in a recall-oriented way, i.e. slightly favouring recall when choosing between models of similar overall performances, while citation deletion was done on neutral and future work functions in a precision-oriented fashion for the purpose of minimising the number of mistakenly deleted citations. Finally, semantic main path networks were extracted using all the above-built networks as follows: for each semantic citation network, a series of network snapshots were created by fixing the start year and sliding the end year of analysis; the top-*K* key-route main paths were extracted from the snapshots and merged into a single network, i.e. the semantic main path network. Two computational linguistics research areas, namely natural language parsing and automatic text summarization, were used to demonstrate the effectiveness of this approach.

Experimental results showed that each semantic main path networks was able to reveal novel topic branches, new important papers of branches, and the development pathways between papers and branches, thus provided complementary views of domain evolution. For example, for large domains such as natural language parsing that were guided by a few seminal studies (like Penn Treebank) and ground-breaking shared tasks, the semantic main path networks were much better at finding these representative works, such as the two early shared tasks on (multilingual) dependency parsing and more future shared tasks on a plethora of topics including semantic dependency parsing, semantic role labelling and dependency parsing of morphologically rich languages, most of which were missed by traditional main path analysis. For automatic text summarization, the semantic main path network approach was able to find an important novel branch about summarization evaluation and the branch about optimization methods for summarization, at the same time preserve and enrich the multi-document summarization, graph-based ranking and sentence fusion/compression branches that were recognised by the traditional approach. Therefore, the convergence of all types of semantic main path networks is believed to represent the topic evolution of a scientific domain more comprehensively. In addition, semantic main path networks achieved much better topic coherence than traditional approaches which ignore citation semantics. In the extracted semantic main path networks, most recognised citation relations were more relevant to, and thus had higher significance for, uncovering the knowledge flow among scientific ideas. Instead, we could find many neutral citations connecting main path papers in the traditional approach. Therefore, we conclude that the semantic main path network approach can discover more pertinent topic branches and uncover more coherent knowledge flow paths.

6. Acknowledgement

Xiaorui Jiang is partially supported by National Office for Philosophy and Social Sciences of China (18ZDA238).

7. REFERENCES

Abu-Jbara, A., Erza, J., & Radev, D. (2013). Purpose and Polarity of Citation: Towards NLP-based Bibliometrics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'13)*, 596–606. https://aclanthology.org/N13-1067

Bae, D.-H., Hwang, S.-M., Kim, S.-W., & Faloutsos, C. (2014). On Constructing Seminal Paper Genealogy. *IEEE Transactions on Cybernetics*, *44*(1), 54–65. https://doi.org/10.1109/TCYB.2013.2246565.

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP'19)*, 3615–3620. https://aclanthology.org/D19-1371

Batagelj, V. (2003). Efficient Algorithms for Citation Network Analysis. Available at: https://arxiv.org/abs/cs/0309023.

Cohan, A., Ammar, W., van Zuylen, M., & Cady, F. (2019). Structural Scaffolds for Citation Intent Classification in Scientific Publications. In *Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'19)*, 3856–3896. https://aclanthology.org/N19-1361

Ding, J., Xiang, T., Ou, Z., Zuo, W., Zhao, R., Lin, C., Zheng, Y., & Liu, B. (2021). Tell Me How to Survey: Literature Review Made Simple with Automatic Reading Path Generation. Available at: https://arxiv.org/abs/2110.06354

Dong, C., & Schäfer, U. (2011). Ensemble-style Self-training on Citation Classification. In *Proceedings of 5th International Joint Conference on Natural Language Processing (IJCNLP'11)*, 623–631. https://aclanthology.org/I11-1070

Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology*, *54*(5), 400–412. https://doi.org/10.1002/asi.10226

Ghosal, T., Tiwary, P., Patton, R., & Stahl, C. (2022). Towards establishing a research lineage via identification of significant citations. *Quantitative Science Studies* (Advance publication). https://doi.org/10.1162/qss_a_00170

Hassan, S.-U., Safder, I., Akram, A., & Kamiran, F. (2018). A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis. *Scientometrics*, *116*, 973–996. https://doi.org/10.1007/s11192-018-2767-x

Hernández-Alvarez, M., Gómez, J.M., & Martínez-Barco, P. (2017). Citation function, polarity and influence classification. *Natural Language Engineering*, *23*(4), 561–588. https://doi.org/10.1017/S1351324916000346

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39–63. https://doi.org/10.1016/0378-8733(89)90017-8

Jha, R., Abu-Jbara, A., Qazvinian, V., & Radev, D.R., (2017). NLP-driven citation analysis for scientometrics. *Natural Language Engineering*, *23*(1), 93–130. https://doi.org/10.1017/S1351324915000443

Jiang, X., Zhu, X., & Chen, J. (2020). Main path analysis on cyclic citation networks. *Journal of the Association for Information Science and Technology, 71*(5), 578–595. https://doi.org/10.1002/asi.24258

Jochim, C., & Schütze, H. (2012). Towards a Generic and Flexible Citation Classifier Based on a Faceted Classification Scheme. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING'12)*, 1343–1358. https://aclanthology.org/C12-1082

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the Evolution of a Scientific Field through Citation Frames. *Transactions of the Association for Computational Linguistic*, *6*, 391–406. https://doi.org/10.1162/tacl_a_00028

Kim, E. H.J., Jeong, Y. K., Kim, Y.H., & Song, M. (2022). Exploring scientific trajectories of a large-scale dataset using topic-integrated path extraction. *Journal of Informetrics*, *16*(1), 101242. https://doi.org/10.1016/j.joi.2021.101242.

Kuan, C.-H. (2020). Regarding weight assignment algorithms of main path analysis and the conversion of arc weights to node weights. *Scientometrics*, *124*, 775–782. https://doi.org/10.1007/s11192-020-03468-8

Lauscher, A., Brandon, K., Kuehl, B., Johnson, S., Jurgens, D., Cohan, A., & Lo, K. (2021). MULTICITE: Modeling realistic citations requires moving beyond the single-sentence single-label setting. *Preprint.* https://arxiv.org/abs/2107.00414

Liu, J. S., & Lu, L. Y. Y. (2012). An integrated approach for main path analysis: The development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, *59*(12), 1948–1962. https://doi.org/10.1002/asi.21692

Liu, J. S., Chen, H.-H., Ho, M. H.-C., & Li, Y.-C. (2014). Citations with different levels of relevancy: Tracing the main paths of legal opinions. *Journal of the American Society of Information Science and Technology*, *65*(12), 2479–2488. https://doi.org/10.1002/asi.23135

Liu, J. S., & Kuan, C.-H. (2016). A new approach for main path analysis: Decay in knowledge diffusion. *Journal of the Association for Information Science and Technology*, *67*(2), 465–476. https://doi.org/10.1002/asi.23384

Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transition in HistCite-based historiograms. *Journal of the American Society of Information Science and Technology*, *27*(1), 25–45. https://doi.org/10.1002/asi.20903

Lyu, D., Ruan, X., Xie, J., & Cheng, Y. (2021). The classification of citing motivations: a meta-synthesis. *Scientometrics*, *126*, 3243–3264. https://doi.org/10.1007/s11192-021-03908-z

Munkhdalai, T., Lalor, J., & Yu, H. (2016). Citation Analysis with Neural Attention Models. In *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI'16)*, 69–77. https://aclanthology.org/W16-6109.

Radev, D. R., Muthukrishnan, P., Qazvinian, V., & Abu-Jbara, A. (2013). The ACL anthology network corpus. *Language Resource and Evaluation*, *47*(4), 919–944. https://doi.org/10.1007/s10579-012-9211-2

Su, X., Prasad, A., Kan, M.-Y., & Sugiyama, K. (2019). Neural Multi-task Learning for Citation Function and Provenance. In *Proceedings of the 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL'19)*, 394–395. https://doi.org/10.1109/JCDL.2019.00122

Tao, S., Wang, X., Huang, W., Chen, W, Wang, T., & Lei, K. (2017). From Citation Network to Study Map: A Novel Model to Reorganize Academic Literatures. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW'17 Companion)*, 1225–1232. https://doi.org/10.1145/3041021.3053059

Teufel, S., Siddharthan, A., & Tidhar, D. (2006a). An annotation scheme for citation function. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue (SIGdial'06)*, 80–87. https://aclanthology.org/W06-1312

Teufel, S., Siddharthan, A., & Tidhar, D. (2006b). Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*, 103–110. https://aclanthology.org/W06-1613

Teufel, S. (2010). The Structure of Scientific Articles: Applications to Citation Indexing and Summarization. Centre for the Study of Language & Information.

Valenzuela, M., Ha, V., & Etzioni, O. (2015). Identifying Meaningful Citations. In Proceedings of the Workshops of Scholarly Big Data: AI Perspectives, Challenges, and Ideas at the 29th AAAI Conference on Artificial Intelligence. https://allenai.org/data/meaningful-citations

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. In *Proceedings of the Thirty-first Conference on Neural Information Processing System (NIPS'17)*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, *10*(1), 93–115. https://doi.org/10.1142/S0219525907000945
Yu, D., & Pan, T. (2021). Tracing the main path of interdisciplinary research considering citation preference: A case from blockchain domain. *Journal of Informetrics*, *16*(2), 101136. https://doi.org/10.1016/j.joi.2021.101136.
Yu, D., & Sheng, L. (2021). Influence difference main path analysis: Evidence from DNA and blockchain domain citation networks. *Journal of Informetrics*, *15*(4), 101186. https://doi.org/10.1016/j.joi.2021.101186.
Zhang, H., Li, L., Li, T., & Wang, D. (2014). PatentDom: Analyzing Patent Relationships on Multi-View Patent Graphs. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM'14)*, 1369–1378. https://doi.org/10.1145/2661829.2662031
Zhuge, H. (2006). Discovery of knowledge flow in science. *Communications of the ACM*, *49*(5), 101–107. https://doi.org/10.1145/1125944.1125948