Extracting the Evolutionary Backbone of a Scientific Domain: The Semantic Main Path Network Approach based on Citation Context Analysis

Supplementary Materials

Xiaorui Jiang[1], Junjun Liu[2]

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

[2] ych1776@163.com

A. No Single Model Beats All

Table S1-S3 show the per-class performances of some selected models (together with the seeds used to train the models). From Table S1, we can see that with the 11-class scheme, the best performing model (variant) 8 (seed = 47353) only won on the "Future" class, while its performance on the "Basis" (="Extends") were significantly poorer than the second best model, i.e. model 4 (47353). By comparison Table S1 and Table S2, the best performance for "Similar" was achieved by model 9 (32491) with the 11-class scheme. However, the best performances for "Background" (="Neutral") was achieved with the 7-class scheme by model 8 (47353) and model 2 (5171). Note that the "Background" class in the 7-class scheme absorbed a small number of "Weakness" instances. As "Weakness" is less relevant to the flow of ideas of scientific innovation, this result is promising for screening out incidental citations (Jochim and Schütze, 2012). While "Uses" (="Usage") received the best performance with the 7-class scheme (model 2, seed = 5171), the best "Extends" (= "Basis") model was model 4 (5171). Similarly, Table S3 shows that model 11 (25603 and 47353) performed the best on "Motivation".

Table S1. Per-Class Performances of Selected Models with the 11-Class Scheme

ID (seed)	2 (25603)			3 (32491)			4 (47353)			8 (47353)			9 (32491)			11 (47353)		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Macro Avg	65.03	64.56	64.26	65.70	61.14	62.52	66.55	64.90	**65.27**	69.05	63.05	**65.39**	67.76	64.60	**65.37**	66.45	63.62	64.62
Future	73.68	82.35	77.78	78.57	64.71	70.97	81.25	76.47	78.79	92.89	76.47	**83.87**	86.67	76.47	81.25	86.67	76.47	81.25
Neutral	76.04	74.74	75.39	76.00	77.82	**76.90**	75.17	77.47	76.30	69.64	79.86	74.40	75.00	75.77	75.38	75.68	76.45	76.06
Similar	65.79	59.52	62.50	57.14	66.67	61.54	59.57	66.67	62.92	60.98	59.52	60.24	62.79	64.29	**63.53**	56.82	59.52	58.14
Motivation	58.02	81.03	67.63	66.13	70.69	**68.33**	72.55	63.79	67.89	62.71	63.79	63.25	54.32	75.86	63.31	54.41	63.79	58.73
Usage	80.45	70.86	75.35	69.88	76.82	73.19	78.17	73.51	**75.77**	77.86	72.19	74.91	77.04	68.87	72.73	72.37	72.85	72.61
Basis	62.92	50.00	55.74	62.50	44.12	51.72	74.07	58.82	**65.57**	71.43	44.12	54.55	54.29	55.88	55.07	63.33	55.88	59.38

Table S2. Per-Class Performances of Selected Models with the 7-Class Scheme

ID (seed)	2 (5171)			3 (5171)			4 (47353)			8 (47353)			9 (25603)			11 (5171)		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Macro Avg	69.73	70.15	69.83	70.62	70.07	70.17	69.31	70.31	69.73	71.60	69.52	**70.29**	70.23	69.81	69.76	74.48	68.30	**70.93**
Future	75.00	70.59	72.73	90.00	69.23	78.26	73.68	82.35	77.78	80.00	70.59	75.00	77.78	82.35	**80.00**	80.00	70.59	75.00
Background	82.17	82.87	**82.52**	80.29	83.90	82.05	80.22	82.02	81.11	79.53	86.24	**82.75**	78.44	81.74	80.06	77.15	88.20	82.31
Similar	62.07	58.06	60.00	63.16	58.06	60.50	59.68	59.68	59.68	59.09	62.90	60.94	66.04	56.45	60.87	64.29	58.06	61.02
Motivation	61.19	70.69	65.60	64.10	56.82	60.24	63.33	65.52	64.41	70.37	65.52	**67.86**	57.75	70.69	63.57	66.67	68.97	**67.80**
Uses	79.43	74.17	**76.71**	75.00	76.32	75.65	79.85	70.86	75.09	83.90	65.56	73.61	82.03	69.54	75.27	83.74	68.21	75.18
Extends	60.53	67.65	63.89	59.09	52.00	55.32	64.71	64.71	**64.71**	57.89	64.71	61.11	64.52	58.82	61.54	72.50	52.94	61.02

Table S3. Per-Class Performances of Selected Models with the 6-Class (`Jurgens2018`) Scheme

ID (seed)	2 (47353)			4 (5171)			8 (47353)			9 (13249)			11 (25603)			11 (47353)		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Macro Avg	73.03	69.17	70.78	75.03	72.29	**73.22**	72.50	70.07	71.17	74.52	72.20	73.00	75.14	72.16	**73.30**	75.38	72.97	**73.99**
Future	76.47	76.47	76.47	81.25	76.47	78.79	81.25	76.47	78.79	77.78	82.35	80.00	82.35	82.35	**82.35**	81.25	76.47	78.79
Background	70.29	82.10	76.77	78.22	78.70	**78.46**	77.54	77.78	77.66	76.07	76.54	76.31	75.00	80.56	77.68	75.37	79.32	77.29
Motivation	71.43	60.43	65.42	59.15	72.41	65.12	70.91	67.24	69.03	66.10	67.24	65.52	69.84	75.86	**72.73**	74.55	70.69	**72.57**
Uses	83.19	65.56	73.33	82.03	69.54	**75.27**	76.09	69.54	72.66	72.05	76.82	74.36	77.69	66.89	71.89	81.97	66.23	73.26
Extends	62.50	58.82	60.61	80.77	61.76	**70.00**	60.00	52.94	56.25	79.17	55.88	65.52	72.00	52.94	61.02	67.65	67.65	67.65

B. Experimental Setup

We performed two case studies to demonstrate the effectiveness of semantic main path network analysis. The underlying dataset was the 2015 version of ACL Anthology Network[1] (AAN; Radev et al., 2013) about computational linguistics papers. The subset between 1985 and 2015 was used because the PDF qualities of papers that are too old are very poor, causing too many parsing errors. The papers were crawled from ACL Anthology[2], the repository hosting papers published in venues sponsored by the Association for Computational Linguistics (ACL). We parsed the papers' PDF documents using by Allen AI's s2orc-doc2json tool[3] and extracted citation contexts. In-text citations were extracted by our own in-house high-precision regular expression-based citation string parser and consecutive citation strings were merged into CITSEG for citation function classification.

The first area of analysis was natural language parsing (parser). Core papers were fetched from AAN by matching the following keywords in paper titles: "parser", "parsing", "parse", "parsed". Then, we added the papers that both cite and are cited by the core papers. The parser citation network, named `AANPar`, was derived from this set of papers. The second area was automatic document summarization (summariser). The semantic citation network, named `AANSum`, was built in a similar way using the following keywords: "summarization", "summarizer", "summarize", "summarized", "summarizing", "summary" and "summaries"[4]. For each area, four semantic citation networks were induced based on adding/deleting in-text citations of certain functions. Their names were ended with `_add_Ext_Mot`, `_plus_add_Use`, `_plus_add_Sim`, and `_del_Bkg_Fut` The start year was fixed to 1985 and the end years were [1995, 2000, 2005, 2010, 2015].

Table S4 summarises the statistics of the resultant semantic citation networks (CITNET) and their largest connected components (CC). We could see when citation networks became larger, the largest CCs became better surrogates of the original citation networks because their sizes got very close.

---

[1] https://aan.how/
[2] https://aclanthology.org/
[3] https://github.com/allenai/s2orc-doc2json
[4] We expected the papers whose title covered "summarisation", "summariser", "summarised" to be induced in the expansion stage.

Parsing		AANPar	add_Ext_Mot	plus_add_Use	plus_add_Sim	del_Bkg_Fut
~1995	CITNET	402/765	105/33	145/64	167/96	334/611
	CC	334/757	7/6	22/21	38/39	314/0605
~2000	CITNET	707/1988	272/142	354/286	391/366	619/1527
	CC	619/1984	48/53	181/233	229/308	588/1519
~2005	CITNET	1091/4188	548/487	671/952	717/1142	984/3131
	CC	984/4182	297/405	487/895	551/1085	941/3122
~2010	CITNET	2017/11210	1278/1780	1488/3578	1546/4155	1882/7961
	CC	1882/11204	954/1680	1273/3515	1367/4099	1810/7949
~2015	CITNET	2823/19212	1898/3212	2205/6618	2279/7583	2629/13413
	CC	2629/19208	1535/3122	1990/6558	2095/7530	2552/13401
Summarisation		AANSum	add_Ext_Mot	plus_add_Use	plus_add_Sim	del_Bkg_Fut
-2000	CITNET	89/85	24/11	31/13	32/13	46/57
	CC	46/84	6/5	6/5	6/5	40/50
-2005	CITNET	222/449	100/69	127/104	132/116	166/282
	CC	166/449	29/30	74/92	78/103	146/277
-2010	CITNET	449/1170	249/215	305/359	317/405	372/706
	CC	372/1170	169/203	231/350	249/399	324/698
-2015	CITNET	640/2113	376/383	460/678	476/769	548/1280
	CC	548/2112	279/368	377/670	397/761	497/1275
Translation		AANMT	add_Ext_Mot	plus_add_Use	plus_add_Sim	del_Bkg_Fut
-1995	CITNET	256/446	53/23	80/30	90/38	180/354
	CC	180/437	11/13	11/13	13/16	169/352
-2000	CITNET	450/1086	140/90	198/153	213/182	341/819
	CC	341/1077	44/63	80/121	96/150	317/808
-2005	CITNET	789/2850	363/321	456/656	475/760	650/2044
	CC	650/2838	186/264	322/619	343/722	616/2031
-2010	CITNET	1742/10528	1068/1518	1273/3700	1298/4166	1541/7083
	CC	1541/10513	815/1454	1102/3657	1136/4122	1484/7072
-2015	CITNET	2842/21390	1943/3068	2292/8048	2322/8979	2608/14273
	CC	2608/21381	1575/2987	2081/7996	2125/8928	2521/14262

C. Main Path Papers for Case Study 1: Natural Language Parsing

This section contains the tables of more complete lists of main path papers extracted from the original semantics-agnostic citation network and the four types of semantic citation networks built based on citation function classification results, with a focus on highlighting their differences.

Table S5. 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**
E91-1035	Proof Figures And Structural Operators For **Categorial Grammar**
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
P89-1010	Word Association Norms **Mutual Information** And Lexicography
H92-1026	Towards History-Based Grammars: Using Richer Models For **Probabilistic Parsing**
H94-1052	Decision Tree Parsing Using A Hidden Derivation Model Excerpt: In this paper, we present a method for contructing a model for the **conditional distribution** of trees given a sentence without the need to define a grammar.
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**.
H93-1047	Automatic Grammar Induction And Parsing Free Text: A **Transformation-Based** Approach
Branch 3	
J93-2004	Building A Large Annotated Corpus Of English: The **<u>Penn Treebank</u>**
P93-1035	Automatic Grammar Induction And Parsing Free Text: A **Transformation-Based** Approach Excerpt: All of the experiments presented below were done using the **<u>Penn Treeban</u>**k annotated corpus (MSM93). (MSM93: J93-2004)
H94-1020	The **<u>Penn Treebank</u>**: 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
P96-1025	New **Statistical Parser** Based On Bigram Lexical Dependencies
C96-1058	Three New **Probabilistic Models** For *Dependency Parsing*: An Exploration
P97-1003	**Three Generative Lexicalized Models For Statistical Parsing**
W98-1118	Exploiting Diverse Knowledge Sources Via **Maximum Entropy** In Named Entity Recognition
A00-2018	A **Maximum-Entropy-Inspired** Parser
C00-1011	Parsing With The Shortest Derivation (**Data-Oriented Parsing**, **DOP** by Rens Bod)
P00-1009	An Improved Parser For **Data-Oriented** Lexical-Functional Analysis (**Data-Oriented Parsing**, **DOP** by Rens Bod)
P02-1034	New **Ranking Algorithms** For Parsing And Tagging: Kernels Over Discrete Structures And The **Voted Perceptron**
W02-1001	**Discriminative Training** Methods For Hidden Markov Models: Theory And Experiments With **Perceptron Algorithms**
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*
W05-1506	Better **K-Best Parsing**

Table S5. Main Path Network Papers of the Parser Network `AANPar` (contd.)

ACLID	Title
Branch 5	
W05-1516	Strictly Lexical **Dependency Parsing**
H05-1066	Non-Projective **Dependency Parsing** Using Spanning Tree Algorithms
P05-1067	Machine Translation Using Probabilistic Synchronous **Dependency** Insertion Grammars
W06-2920	**CoNLL-X Shared Task On Multilingual Dependency Parsing**
W06-2932	**Multilingual Dependency** Analysis With A Two-Stage Discriminative Parser
D07-1013	Characterizing the Errors of Data-Driven **Dependency Parsing** Models
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
W07-2220	Data-Driven **Dependency Parsing** across Languages and Domains: Perspectives from the CoNLL-2007 Shared task
Branch 6	
W10-1401	Statistical Parsing of **Morphologically Rich** Languages (SPMRL) What How and Whither
W10-1408	Handling Unknown Words in Statistical Latent-Variable Parsing Models for **Arabic English** and **French**
W10-1410	Lemmatization and Lexicalized Statistical Parsing of **Morphologically-Rich** Languages: the Case of **French**
Branch 7	
P10-1040	Word Representations: A Simple and General Method for **Semi-Supervised Learning**
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

Table S6. Main Path Network Papers the Parser Network `AANPar_add_Ext_Mot`

ACLID	Title
Branch 1	
P85-1018	Using Restriction To Extend Parsing Algorithms For **Complex-Feature-Based Formalisms**
J89-4001	A Parsing Algorithm For **Unification Grammar**
E91-1031	Prediction In **Chart Parsing** Algorithms For **Categorial Unification Grammar**
E89-1032	An Algorithm For Generation In **Unification Categorial Grammar**
P85-1021	Parsing **Head-Driven Phrase Structure Grammar**
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 3	
H89-2014	Augmenting A **Hidden Markov Model** For Phrase-Dependent Word Tagging
P90-1031	Parsing The **LOB Corpus**
J93-2006	Coping With Ambiguity And Unknown Words Through **Probabilistic Models**
Branch 4	
W96-0213	A **Maximum Entropy** Model for Part-of-speech Tagging
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
H05-1100	Morphology And **Reranking** For The Statistical Parsing Of Spanish
W97-0302	Global Thresholding And **Multiple-Pass Parsing**
P05-1022	Coarse-To-Fine **N-Best Parsing** And MaxEnt **Discriminative Reranking**
W05-1506	Better **K-Best Parsing**
W05-1505	**Corrective Modeling** for Non-Projective Dependency Parsing Explanation: The technique proposed in this paper is similar to that of recent parser reranking approaches (Collins, 2000; Charniak and Johnson, 2005); … (an ICML conference paper extended to J05-1003 with the same title; P05-1022)
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** Lexicalized Models For **Statistical Parsing**
A00-2018	A **Maximum-Entropy-Inspired** Parser
W00-1201	Two Statistical Parsing Models Applied To The ***Chinese Treebank***
C02-1126	Recovering Latent Information in Treebanks
J04-4004	Intricacies Of **Collins Parsing** Model (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 Excerpt: 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?

Table S6. Main Path Network Papers the Parser Network `AANPar_add_Ext_Mot` (contd.)

ACLID	Title
Branch 7	
W04-0307	A Statistical Constraint Dependency Grammar (CDG) Parser
H05-1066	Non-Projective **Dependency Parsing** Using Spanning Tree Algorithms
W06-2932	**Multilingual Dependency Analysis** With A Two-Stage Discriminative Parser
D07-1013	Characterizing the Errors of Data-Driven **Dependency Parsing** Models
D07-1096	**The CoNLL 2007 Shared Task for Dependency Parsing**
D07-1112	Frustratingly Hard Domain Adaptation for **Dependency Parsing**
P10-1002	**Dependency Parsing** and Projection Based on Word-Pair Classification
D08-1059	A Tale of Two Parsers: Investigating and Combining Graph-based and Transition-based **Dependency Parsing**
P07-1050	K-best Spanning Tree Parsing
P11-2033	Transition-based **Dependency Parsing** with Rich Non-local Features
D12-1133	Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective **Dependency Parsing**
C14-1078	Feature Embedding for **Dependency Parsing**
C14-1051	Jointly or Separately: Which is Better for Parsing Heterogeneous **Dependencies**?
P14-2107	Enforcing Structural Diversity in Cube-pruned **Dependency Parsing**
Branch 8	(A "wrong" branch of machine translation using dependency syntax)
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
P05-1012	*Online Large-Margin Training* Of **Dependency Parsers**

Table S7. Main Path Network Papers the Parser Network `AANPar_plus_add_Use`

ACLID	Title
Branch 1	
A88-1019	A **Stochastic** Parts Program And Noun Phrase **Parser** For Unrestricted Text
P89-1010	Word Association Norms **Mutual Information** And Lexicography
J93-1007	Retrieving Collocations From Text: Xtract
P93-1024	**Distributional** Clustering Of English Words
H91-1046	A Trellis-Based Algorithm For Estimating The Parameters Of **Hidden Stochastic Context-Free Grammar**
J93-1002	Generalized **Probabilistic LR Parsing** Of Natural Language (**Corpora**) With Unification-Based Grammars
J90-1003	Word Association Norms **Mutual Information** And Lexicography
P90-1034	Noun Classification From Predicate-Argument Structures
C90-3010	Acquisition Of Lexical Information From A Large Textual Italian **Corpus**
J93-2005	Lexical Semantic Techniques For **Corpus** Analysis
J93-1001	**Introduction To The Special Issue On Computational Linguistics Using Large Corpora**
H89-2014	Augmenting A **Hidden Markov Model** For Phrase-Dependent Word Tagging
P90-1031	Parsing The LOB **Corpus**
J93-2006	Coping With Ambiguity And Unknown Words Through **Probabilistic Models**
Branch 2	
H93-1047	Automatic Grammar Induction and Parsing Free Text: A *Transformation-Based* Approach
W95-0107	**Text Chunking** using *Transformation-Based* Learning
P98-1010	A Memory-Based Approach to Learning **Shallow** Natural Language Patterns
W99-0621	A Learning Approach to **Shallow Parsing**
	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.).
W00-0721	**Shallow Parsing** by Inferencing with Classifiers
Branch 3	(multiple intelligent sources of statistical parsing techniques)
E85-1024	A **Probabilistic Parser**
J94-2001	Tagging English Text With A **Probabilistic Model**
J93-2004	**Building A Large Annotated Corpus Of English: The Penn Treebank**
W96-0213	A **Maximum Entropy Model** for Part-of-speech Tagging
C02-1126	Recovering Latent Information In **Treebanks**
P00-1058	**Statistical Parsing** With An Automatically-Extracted Tree Adjoining Grammar
W00-1201	Two **Statistical Parsing** Models Applied To The ***Chinese Treebank***
J04-4004	Intricacies Of **Collins Parsing Model**
J95-4004	**Transformation-Based-Error-Driven** Learning And Natural Language Processing: A Case Study In Part-Of-Speech Tagging
P01-1067	A Syntax-Based *Statistical Translation Model*
N04-1013	Speed And Accuracy In Shallow And Deep **Stochastic Parsing**
W04-3223	Incremental Feature Selection And L1 Regularization For Relaxed **Maximum-Entropy Modeling**
J05-1003	**Discriminative Reranking** For Natural Language Parsing
H05-1100	Morphology And **Reranking** For The **Statistical Parsing** Of Spanish
Branch 4	
W04-0814	The University Of Amsterdam At Senseval-3: **Semantic Roles** And Logic Forms
W05-0624	Sparse Bayesian Classification Of **Predicate Arguments**
W05-0620	**Introduction To The CoNLL-2005 Shared Task: Semantic Role Labeling**
Branch 5	(About dependency parsing similar to Branch 7 in Table S6)
…	
Branch 6	(About semi-supervised learning and cross-lingual dependency parsing extended from Branch 7 in Table S6)
P08-1068	Simple **Semi-supervised** Dependency Parsing
D09-1087	**Self-Training** PCFG Grammars with Latent Annotations Across Languages
D09-1058	An Empirical Study of **Semi-supervised** Structured Conditional Models for Dependency Parsing
P10-1040	Word Representations: A Simple and General Method for **Semi-Supervised Learning**
N12-1052	**Cross-lingual** Word Clusters for Direct **Transfer** of Linguistic Structure
D12-1001	Syntactic **Transfer** Using a **Bilingual** Lexicon
W14-1613	Distributed Word Representation Learning for **Cross-Lingual** Dependency Parsing
W14-1614	Treebank Translation for **Cross-Lingual** Parser Induction

W14-4202	Learning from a Neighbor: Adapting a Japanese Parser for Korean Through Feature **Transfer Learning**

Table S8. Main Path Network Papers the Parser Network `AANPar_plus_add_Sim`

ACLID	**Title**
Branch 1	
P99-1047	A Decision-Based Approach To **Rhetorical Parsing**
A00-2002	The Automatic Translation Of **Discourse Structures**
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**
C04-1041	The Importance Of Supertagging For Wide-Coverage **CCG Parsing**
C04-1180	Wide-Coverage Semantic Representations From A **CCG Parser**
W04-3215	Object-Extraction And Question-Parsing Using **CCG**
Branch 3	
P05-1012	Online Large-Margin Training Of **Dependency Parsers**
H05-1066	Non-Projective **Dependency Parsing** Using Spanning Tree Algorithms
N06-2033	Parser Combination By **Reparsing**
P05-1013	Pseudo-Projective **Dependency Parsing**
W06-2920	**CoNLL-X Shared Task On Multilingual Dependency Parsing**
W06-2932	**Multilingual Dependency** Analysis With A Two-Stage Discriminative Parser
D07-1013	Characterizing the Errors of Data-Driven **Dependency Parsing** Models
N07-1050	Incremental Non-Projective **Dependency Parsing**
D07-1097	Single Malt or Blended? A Study in **Multilingual Parser** Optimization
D07-1096	**The CoNLL 2007 Shared Task on Dependency Parsing**
D07-1101	Experiments with a Higher-Order Projective **Dependency Parser**
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	
P08-1068	Simple **Semi-supervised** Dependency Parsing
D09-1087	**Self-Training** PCFG Grammars with Latent Annotations **Across Languages**
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
…	(The remaining same as Branch 6 in Table S7)

Table S9. Main Path Network Papers the Parser Network `AANPar_del_Bkg_Fut`

ACLID	Title
Branch 1	
P02-1034	New Ranking Algorithms For Parsing And Tagging: **Kernels** Over Discrete Structures And The **Voted Perceptron**
W02-1001	Discriminative Training Methods For **Hidden Markov Models**: Theory And Experiments With **Perceptron Algorithms**
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 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
N12-1052	**Cross-lingual** Word Clusters for Direct **Transfer** of Linguistic Structure
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	The LIGM-Alpage architecture for the **SPMRL 2013 Shared Task**: Multiword Expression Analysis and Dependency Parsing
W13-4906	Exploring beam-based shift-reduce dependency parsing with DyALog: Results from the **SPMRL 2013 shared task**
W13-4910	**SPMRL'13 Shared Task** System: The CADIM Arabic Dependency Parser

D. Citation Context Excerpts of Case Study 1: Natural Language Parsing

This section presents more citation context excepts to help readers understand Figure 8-11 in the main text.

D.1 Main Path Network

Nothing really.

D.2 Main Path Network: Add Extension and Motivation Citations

On the semantic main path network extracted by adding extension and motivation citations (Figure 9 in the main text), we should expect misrecognised or vague extension/motivation cases because they were more difficult classes. An example was the citation from P03-1055 ("Deep Syntactic Processing by Combining Shallow Methods") to E03-1055 ("An efficient implementation of a new DOP model"). From the citation context excerpt below, it looks more appropriate to treat it as a weakness citation (about research gap). It was recognised as "Motivation" because our refined annotation guidelines allow the fact that research gap exists or problem is challenging to motive current study, similar to Teufel's rule about "problem worth studying" (Teufel, 2010). Indeed, the citing paper explored "a novel approach for finding long-distance dependencies" (in its Abstract) to solve this hard problem.

> *"Clearly, information about long-distance relationships is vital for semantic interpretation. However, such constructions prove to be difficult for stochastic parsers (Collins et al., 1999) and they either avoid tackling the problem (Charniak, 2000; <u>Bod, 2003</u>) or only deal with a subset of the problematic cases (Collins, 1997)."*

Below shows more citation context excerpts corresponding to the citations in Figure 9 in the main text that are annotated with citation functions.

W96-0213 → J94-2001 | H94-1052 (Extension): *"Most of the recent corpus-based POS taggers in the literature are either statistically based, and use Markov Model (Weischedel et al., 1993, <u>Merialdo, 1994</u>) or Statistical Decision Tree (Jelinek et al., 1994, Magerman, 1995) (SDT) techniques, or are primarily rule based, such as Drill's Transformation Based Learner (Drill, 1994) (TBL). The Maximum Entropy (MaxEnt) tagger presented in this paper combines the advantages of all these methods. It uses a rich feature representation, like TBL and SDT, and generates a tag probability distribution for each word, like Decision Tree and Markov Model techniques."*

H05-1066 → P05-1012 (Extension): *"Using this spanning tree representation, we extend the work of McDonald et al. (2005) on online large-margin discriminative training methods to non-projective dependencies."*

H05-1066 → W04-0307 (Motivation): *"However, non-projective analyses have recently attracted some interest, not only for languages with freer word order but also for English. In particular, Wang and Harper (2004) describe a broad coverage non projective parser for English based on a hand-constructed constraint dependency grammar rich in lexical and syntactic information.*

D08-1059 → P07-1050: (Motivation): *"Our combined parser makes the biggest contribution of this paper. In contrast to the models above, it includes both graph-based and transition-based components. An existing method to combine multiple parsing algorithms is the ensemble approach (Sagae and Lavie, 2006a), which was reported to be useful in improving dependency parsing (Hall et al., 2007)."*

D08-1059 → D07-1013 (Motivation): *"McDonald and Nivre (2007) showed that the MSTParser and MaltParser produce different errors. This observation suggests a combined approach: by using both graph-based information and transition-based information, parsing accuracy can be improved."*

D.3 Main Path Network: Further Add Usage Citations

On the semantic main path network extracted by further adding usage citations (Figure 10 in the main text), we present some additional excerpts of usage citation contexts as follows.

W00-1201 → P00-1058 (Usage): *"A different rule is used for extracting auxiliary trees; see (Chiang, 2000) for details."*

W00-1201 → P00-1058 (Usage): *"For our model we break down these probabilities further: first the elementary tree is generated without its anchor, and then its anchor is generated. See (Chiang, 2000) for more details."* (According to Teufel's annotation guidelines, the steps used in the cited paper are used by the citing paper, so this is a usage citation.)

P00-1058 → W96-0213 (Usage): *"Following (Collins, 1997), words occurring fewer than four times in training were replaced with the symbol *UNKNOWN* and tagged with the output of the part-of-speech tagger described in (Ratnaparkhi, 1996)."*

W96-0213 → J93-2004 (Usage): "The experiments in this paper were conducted on … Penn Treebank project (Marcus et al., 1994)," (typo in paper, but matching correct)

P00-1058 → W97-0302 (Usage): *"We use a beam search, computing the score of an item [η; i; j] by multiplying it by the prior probability P (η) (Goodman, 1997); …"*

W97-0302 → P96-1030: (Extends): *"The key insight of global thresholding is due to Rayner and Carter (1996)."*

P96-1030 → A92-1018 (Extends): *"The work reported here is a logical continuation of two specific strands of research aimed in this general direction. The first is the popular idea of statistical tagging e.g. (DeRose, 1988; Cutting et al., 1992; Church, 1988)."*

P01-1067 → J95-4004 (Usage): "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."

D.4 Main Path Network: Further Add Similarity Citations

On the semantic main path network extracted by further adding similarity citations (Figure 11 in the main text), we could find a borderline case between "Similar", "Usage" and even "Extension". The citation was from C02-1126 to W00-1201. From the citation context excerpts (see below), the "with the same head rules" might imply similarity. However, the whole sentence looks to mean that the cited paper's rules were used in "the third experiment". From the next context sentence, "*although we have modified them for parsing*" seems to imply an extension (technical modification) from the cited paper. This example well shows the vagueness between the usage class and similarity class.

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

Below show two more examples of blurred border between the similarity class and usage class.

W03-1005 → P03-1055 (Similar? Usage?) : *"In the experiments we use the same training, test, and development data as in Dienes and Dubey (2003), where non-local dependencies are annotated with the help of empty elements (EEs) co-indexed with their controlling constituents (if any)."*

D09-1058 → P08-1068 (Similar? Usage?): *"Since this method only considers projective dependency structures, we "projectivized" the PDT training data in the same way as (Koo et al., 2008)."*

We present some additional excerpts of similarity citation contexts below.

W97-0301 → J96-1002 (Similar): *"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)."*

P08-1068 → D07-1101 (Similar): *"The feature sets we used are similar to other feature sets in the literature (McDonald et al., 2005a; Carreras, 2007)."*

N12-1052 → P10-1040 (Similar): *"The feature model used for the NER tagger is shown in Table 2. These are similar to the features used by Turian et al. (2010), with the main difference that we do not use any long range features and that we add templates that conjoin adjacent clusters and adjacent tags as well as templates that conjoin label transitions with tags, clusters and capitalization features."* (This example can also be regarded as a "CoCoGM", therefore justify the possibility of multi-label citation function classification)

D.5 Main Path Network: Delete Neutral and Future Work Citations

On the semantic main path network extracted by removing neutral and future work citations (Figure 12 in the main text), we present additional citation context excerpts below.

D07-1119 → N07-1049 (Usage): *"We then used a parsing revision technique (Attardi and Ciaramita, 2007) to learn how to correct these errors, producing a parse reviser called DesrReviser."*

D07-1096 → D07-1119 (Usage?): *"Transition-based parsers either maintain a classifier that predicts the next transition or a global probabilistic model that scores a complete parse. To train these classifiers and probabilitistic models several approaches were used: SVMs (Duan et al., 2007; Hall et al., 2007a; Sagae and Tsujii, 2007), modified finite Newton SVMs (Wu et al., 2007), maximum entropy models (Sagae and Tsujii, 2007), multiclass averaged perceptron (Attardi et al., 2007) and maximum likelihood estimation (Watson and Briscoe, 2007)."*

E. Main Path Papers for Case Study: Automatic Text Summarisation

This section contains the tables of more complete lists of main path papers extracted from the original semantics-agnostic citation network and the four types of semantic citation networks built based on citation function classification results.

Table S10. Main Path Network Papers of the Summarisation Network `AANSum`

ACLID	Title
Branch 1	
C86-1061	Tailoring Importance Evaluation To Reader's Goals: A Contribution To Descriptive Text Summarization
C94-1056	Abstract Generation Based On **Rhetorical Structure** Extraction
W97-0713	From **Discourse Structures** To Text Summaries
W97-0703	Using **Lexical Chains** For Text Summarization
Branch 2	
J98-3005	Generating Natural Language Summaries From **Multiple On-Line Sources**
C00-1024	A Muitilingual News Summarizer
P99-1071	Information Fusion In The Context Of **Multi-Document Summarization**
C00-2129	Multi-Topic **Multi-Document Summarization**
W00-0403	**Centroid-Based** Summarization Of **Multiple Documents**: Sentence Extraction Utility-Based Evaluation And User Studies
W00-0404	Extracting Key Paragraph Based On Topic And Event Detection Towards **Multi-Document Summarization**
H01-1065	Sentence Ordering In **Multidocument Summarization**
H01-1054	**Multidocument Summarization** Via Information Extraction
W00-1009	A Common Theory Of Information Fusion From Multiple Text Sources Step One: **Cross-Document Structure**
H01-1056	NewsInEssence: A System For Domain-Independent Real-Time News Clustering And **Multi-Document Summarization**
W03-0510	The Potential And Limitations Of Automatic Sentence Extraction For Summarization
H01-1065	Sentence Ordering In **Multidocument Summarization**
W02-0402	Selecting Sentences For **Multidocument Summaries** Using Randomized Local Search
J02-4001	**Introduction To The Special Issue On Summarization**
Evaluation	
N03-1020	**Automatic Evaluation** Of Summaries Using N-Gram Co-Occurrence Statistics
Branch 3	
P04-3020	**Graph-Based Ranking** Algorithms For Sentence Extraction Applied To Text Summarization
W04-3252	**TextRank**: Bringing Order Into Texts
P05-3013	Language Independent Extractive Summarization Excerpt: We demonstrate **TextRank**
I05-2004	A Language Independent Algorithm for Single and **Multiple Document Summarization** Explanation: relies on iterative **graph-based ranking** algorithm
Branch 4	
P10-1058	Automatic Generation of Story Highlights joint content selection and **compression** model for single-document summarization
P11-1049	Jointly Learning to Extract and **Compress** Excerpt: Inference in our model can be cast as an **ILP** and thereby solved in reasonable time
E12-1023	Large-Margin Learning of **Submodular** Summarization Models
D12-1022	Multiple Aspect Summarization Using **Integer Linear Programming**
P13-1020	Fast and Robust **Compressive** Summarization with **Dual Decomposition** and Multi-Task Learning
D13-1047	Document Summarization via Guided Sentence **Compression**
D13-1156	Fast Joint **Compression** and Summarization via Graph Cuts
P14-2052	Single Document Summarization based on **Nested Tree Structure** Excerpt: … combining sentence selection and sentence **compression** … We used both dependency between words and dependency between sentences by constructing a nested tree, … We formulated a summarization task as a **combinatorial optimization problem**
D14-1196	**Dependency-based Discourse** Parser for Single-Document Summarization

Table S11. Main Path Network Papers the Summarisation Network `AANSum_add_Ext_Mot`

ACLID	Title
Branch 1	
W97-0703	Using **Lexical Chains** For Text Summarization
C00-2140	DIASUMM: Flexible Summarization Of Spontaneous Dialogues In Unrestricted Domains
W97-0713	From **Discourse Structures** To Text Summaries
W98-1124	Improving Summarization Through **Rhetorical** Parsing Tuning
W00-0401	Concept Identification And Presentation In The Context Of Technical Text Summarization
P99-1072	Improving Summaries By Revising Them
	Excerpt: This paper describes a program which revises a draft text by aggregating together descriptions of **discourse entities**, in addition to deleting extraneous information.
J02-4005	Generating Indicative-Informative Summaries With SumUM
J02-4002	Summarizing **Scientific** Articles: Experiments With Relevance And **Rhetorical Status**
C10-2049	Towards Automated **Related Work** Summarization
W03-0505	Summarising Legal Texts: Sentential Tense And **Argumentative Roles**
W04-1007	A **Rhetorical Status** Classifier For Legal Text Summarisation
Branch 2	
J98-3005	Generating Natural Language Summaries From **Multiple On-Line Sources**
W04-3256	**Multi-Document** Biography **Summarization**
Branch 3	
W00-0408	A Comparison Of Rankings Produced By Summarization **Evaluation Measures**
W02-0406	Manual And **Automatic Evaluation** Of Summaries
N03-1020	**Automatic Evaluation** Of Summaries Using N-Gram Co-Occurrence Statistics
W04-1013	ROUGE: A Package For **Automatic Evaluation** Of Summaries
P08-1094	Can You Summarize This? Identifying Correlates of Input Difficulty for Multi-Document Summarization
E09-1062	Performance Confidence Estimation for Automatic Summarization
W11-0504	Who wrote What Where: Analyzing the content of human and automatic summaries
Branch 4-1	
A00-1043	**Sentence Reduction** For Automatic Text Summarization
A00-2024	Cut And Paste Based Text Summarization
	Excerpt: 4.2 Sentence **reduction** … 4.3 Sentence **combination** …
J05-3002	**Sentence Fusion** For **Multidocument News Summarization**
C10-2049	Towards Automated Related Work Summarization (Citation context parsing error)
Branch 4-2	
P08-2049	Query-based **Sentence Fusion** is Better Defined and Leads to More Preferred Results than Generic Sentence Fusion
D08-1019	**Sentence Fusion** via Dependency Graph **Compression**
C10-1037	**Multi-Sentence Compression**: Finding Shortest Paths in Word Graphs
N13-1030	Keyphrase Extraction for N-best Reranking in **Multi-Sentence Compression**
W14-4407	A Template-based **Abstractive** Meeting Summarization: Leveraging Summary and Source Text Relationships
W13-2117	**Abstractive** Meeting Summarization with Entailment and **Fusion**
Branch 5	
W04-3252	**TextRank**: Bringing Order Into Texts
P07-1070	Towards an **Iterative Reinforcement** Approach for Simultaneous Document Summarization and Keyword Extraction
P09-1023	Summarizing Definition from Wikipedia
	Excerpt: Following the **iterative reinforcement approach** for summarization and keyword extraction (Wan et al., 2007), it could be used to refine other versions of summaries. Hence, we …

Table S12. Main Path Network Papers the Summarisation Network `AANSum_plus_add_Use`

ACLID	Title
Branch 1-1	(About **sentence fusion/compression** same as Branch 4-2 in Table S11)
…	
Branch 1-2	
P11-1049	Jointly Learning to Extract and **Compress**
P13-1020	Fast and Robust **Compressive** Summarization with Dual Decomposition and Multi-Task Learning
D13-1047	Document Summarization via Guided **Sentence Compression**
D13-1156	Fast Joint **Compression** and Summarization via Graph Cuts
Branch 2	
W00-0408	A Comparison Of Rankings Produced By Summarization **Evaluation Measures**
W02-0406	Manual And **Automatic Evaluation** Of Summaries
N03-1020	**Automatic Evaluation** Of Summaries Using N-Gram Co-Occurrence Statistics
W04-1013	**ROUGE**: A Package For **Automatic Evaluation** Of Summaries
P04-1027	An Empirical Study Of Information Synthesis Task
P05-1035	QARLA: A Framework For The **Evaluation** Of Text Summarization Systems
W05-0906	**Evaluating** Summaries And Answers: Two Sides Of The Same Coin?
N04-1019	**Evaluating** Content Selection In Summarization: The **Pyramid** Method
W03-0508	**Examining** The Consensus Between Human Summaries: Initial Experiments With **Factoid Analysis**
W04-3254	**Evaluating** Information Content By **Factoid Analysis**: Human Annotation And Stability
W05-0901	A Methodology For **Extrinsic Evaluation** Of Text Summarization: Does **ROUGE** Correlate?
Branch 3	(multiple intelligent sources of statistical parsing techniques)
J02-4002	Summarizing **Scientific** Articles: Experiments With Relevance And ***Rhetorical Status***
W03-0505	Summarising Legal Texts: Sentential Tense And ***Argumentative Roles***
W06-1613	Automatic Classification Of **Citation Function**
N07-1040	Whose Idea Was This and Why Does it Matter? Attributing **Scientific Work** to **Citations**
C08-1087	**Scientific Paper Summarization** Using **Citation** Summary Networks
N09-1066	Using **Citations** to Generate **Surveys of Scientific Paradigms**
C10-1101	**Citation Summarization** Through Keyphrase Extraction
P10-1057	Identifying Non-Explicit **Citing Sentences** for **Citation-Based Summarization**.

Table S13. Main Path Network Papers the Summarisation Network `AANSum_plus_add_Sim`

ACLID	Title
Branch 1	(expanded from Branch 2 about Evaluation in Table S12)
P04-1077	**Automatic Evaluation** Of Machine Translation Quality Using Longest Common Subsequence And Skip-Bigram Statistics
H05-1019	Kernel-Based Approach For **Automatic Evaluation** Of Natural Language Generation Technologies: Application To Automatic Summarization
P06-2020	Topic-Focused Multi-Document Summarization Using An **Approximate Oracle Score**
D09-1032	Automatically **Evaluating** Content Selection in Summarization **without Human Models**
C10-2122	Multilingual Summarization Evaluation **without Human Models**

Table S14. Main Path Network Papers the Summarisation Network `AANSum_Del_Bkg_Fut`

ACLID	Title
Branch 1	
C00-2129	Multi-Topic **Multi-Document Summarization**
J98-3005	Generating Natural Language Summaries From **Multiple On-Line Sources**
P99-1071	Information Fusion In The Context Of **Multi-Document Summarization**
W00-0403	Centroid-Based Summarization Of **Multiple Documents**: **Sentence Extraction** Utility-Based Evaluation And User Studies
H01-1065	Sentence Ordering In **Multidocument Summarization**
H01-1054	**Multidocument Summarization** Via Information **Extraction**
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	
P04-3020	**Graph-Based Ranking** Algorithms For Sentence Extraction Applied To Text Summarization
W04-3252	**TextRank**: Bringing Order Into Texts
I05-2004	A Language Independent Algorithm for Single and Multiple Document Summarization Explanation: Towards an **Iterative Reinforcement Approach** for Simultaneous Document Summarization and Keyword Extraction
P07-1070	Towards an Iterative Reinforcement Approach for Simultaneous Document Summarization and Keyword Extraction Excerpt: Following the **iterative reinforcement approach** for summarization and keyword extraction (Wan et al., 2007), it could be used to refine other versions of summaries. Hence, we …
P08-1041	Summarizing Emails with Conversational Cohesion and Subjectivity Abstract: We first build a sentence quotation graph that captures the conversation structure among emails. … 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 Excerpt: 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 Excerpt (same authors as C10-2105): our method significantly outperforms competitive baselines based on the maximum coverage of important words using **integer linear programming**

F. Citation Context Excerpts of Case Study 2: Automatic Text Summarisation

F.1 Main Path Network

We present here some citation context excerpts to help understand the semantic main path network extracted using the traditional key-route main path approach (Figure 13 in the main text). We see that many citations between main path papers are actually about comparisons.

D07-1047 → I05-2004 (Comparison): *"Although some recent systems indicate an improvement over the baseline (Mihalcea, 2005; Mihalcea and Tarau, 2005), statistical significance has not been shown. We show that by using a neural network ranking algorithm and third-party datasets to enhance sentence features, our system, NetSum, can outperform the baseline with statistical significance."*

P14-2052 → D13-1156 (Weakness or Comparison): *"Extracting a subtree from the dependency tree of words is one approach to sentence compression (Tomita et al., 2009; Qian and Liu, 2013; Morita et al., 2013; Gillick and Favre, 2009). However, these studies have only extracted rooted subtrees from sentences. We allowed our model to extract a subtree that did not include the root word …"* (Another example of multi-label classification. According to our annotation guidelines, the overwriting rule should apply here to recognise it as comparison.)

E12-1023 → P11-1049 (Comparison): *"Another closely related work also takes a maxmargin discriminative learning approach in the structural SVM framework (Berg-Kirkpatrick et al., 2011) … However, they do not consider submodular functions, but instead …"*

D12-1022 → P11-1049 (Comparison: CoCoGM): *"Our work differs from Gillick et al. (2009) and Berg-Kirkpatrick et al. (2011) in three important respects"*

D12-1022 → P11-1049 (Comparison: CoCoGM): *"Thirdly, unlike Berg-Kirkpatrick et al. (2011) our model does not try to learn all the parameters (e.g., content, rewrite rules, style) of the summarization problem jointly; …, the learning process is more robust and reliable."*

D12-1022 → P11-1049 (Comparison: CoCoRes): *"We also compared against the "learned phrase compression" system of Berg-Kirkpatrick et al. (2011) …"*

P11-1049 → P10-1058: (Comparison): *"A second contribution of the current work is to show a system for jointly learning to jointly compress and extract that exhibits gains in both ROUGE and content metrics over purely extractive systems. Both Martins and Smith (2009) and Woodsend and Lapata (2010) build models that jointly extract and compress, but learn scores for sentences (or phrases) using independent classifiers."*

P10-1058 → D07-1047 (Usage): *"Svore et al. (2007) were the first to foreground the highlight generation task which we adopt as an evaluation testbed for our model. Their approach is however a purely extractive one."*

N09-1041 → N06-2046 (Neutral): *"There are several approaches to modeling document content: simple word frequency-based methods (Luhn, 1958; Nenkova and Vanderwende, 2005), graph-based approaches (Radev, 2004; Wan and Yang, 2006), as well as more linguistically motivated techniques (Mckeown et al., 1999; Leskovec et al., 2005; Harabagiu et al., 2007). Another strand of work (Barzilay and Lee, 2004; Daume III and Marcu, 2006; Eisenstein and Barzilay, 2008), has explored the use of structured probabilistic topic models to represent document content. However, little has been done to directly compare the benefit of complex content models to simpler surface ones for generic multi-document summarization"*.

F.2 Semantic Main Path Network: Add Extension and Motivation Citations

On the semantic main path network extracted by adding extension and motivation citations (Figure 14 in the main text), we present additional citation context excerpts of the annotated citation edges below.

J98-3005 → A97-1013 (Extension?): *"We present a system, called SUMMONS 1 (McKeown and Radev 1995; Radev 1996; Radev and McKeown 1997), shown in Figure 1, which introduces novel techniques in the following areas: ..."* (Vague case: This may be an extension because it is a journal extension to previous conference publications with several novel techniques, or this may be just a neutral citations because it is the system presented by the citing paper that "introduced novel techniques in the following areas: …", or this may also be a usage citation because the fact that the current paper present the system of the cited paper indicate a usage?)

W03-0505 → J02-4002 (Extension): *"Our methodology builds and extends the Teufel and Moens (Teufel and Moens, 2002) approach to automatic summarization."*

C00-2140 → W97-0703 (Extension): *"Following (Boguraev and Kennedy, 1997; Barzilay and Elhadad, 1997) who use TextTiling (Hearst, 1997) for their summarization systems of written text, we adapted this algorithm (its block comparison version) for speech data."*

W03-0505 → J02-4002 (Extension): *"Our methodology builds and extends the Teufel and Moens (Teufel and Moens, 2002) approach to automatic summarization."*

P08-1094 → W04-1013 (Motivation subsumes Usage): In a footnote *"[1]The routinely used tool for automatic evaluation ROUGE was adopted exactly because it was demonstrated it is highly correlated with the manual DUC coverage scores (Lin and Hovy, 2003a; Lin, 2004)."* (According to Teufel's guidelines, this citation should be annotated to a motivation citation, because the causal clause explains the positive aspect of ROUGE which motivates the use of ROUGE in this paper. Motivation subsumes usage, although we saw many similar cases were annotated to usage in Teufel's original dataset. We believe this citation was classified to motivation because our dataset has several similar cases where we overwrote usage with motivation in such cases.).

P09-1023 → P08-1094 (Motivation): *"Analyzing the features used could let us understand summarization better (Nenkova and Louis, 2008). Here, we focus on the statistical analysis of …"*

F.3 Semantic Main Path Network: Further Add Usage Citations

On the semantic main path network extracted by further adding usage citations (Figure 15 in the main text), we present some additional citation context excerpts of the annotated citation edges below.

W04-1016 → W03-0508 (Usage): *"Given this data, we first segment the original pair of sentences into "factoids" in the style of Halteren and Teufel (2003)."*

W04-1016 → N04-1019 (Usage): *"The third evaluation we perform ourselves, due to its difficulty. This follows the general rubric described by Nenkova and Passonneau's (2004) pyramid scoring scheme, though it differs in the sense that we …"*

W04-1016 → N04-1019 (Extension): *"This follows the general rubric described by Nenkova and Passonneau's (2004) pyramid scoring scheme, though it differs in the sense that we base our evaluation not on a reference summary, but on the original two document sentences. Our methodology is described below. …"*

N07-1040 → W06-1613 (Usage): *"Another difference is that we use around 1700 additional cue phrases acquired from previous work on another discourse task4 (Teufel et al., 2006)."*

Here is an example of misclassification misguided by the expression "using".

D07-1040 → W04-3247 (Usage, X): *"This similarity is computed as a cosine similarity and by default the sentences that exhibit a cosine similarity higher than 0.7 are not added to the summary. Note that although the MEAD distribution also includes an optional feature calculated* ***using*** *the LexRank graph-based algorithm (Erkan and Radev, 2004), this feature could not be used since it takes days to compute for very long documents such as ours, and thus its application was not tractable."*

F.4 Semantic Main Path Network: Further Add Similarity Citations

On the semantic main path network extracted by further adding similarity citations (Figure 16 in the main text), we present some additional citation context excerpts of the annotated citation edges below.

N04-1019 → W03-0508 (Similar): *"Though we look at multidocument summaries rather than single document ones, SCU annotation otherwise resembles the annotation of factoids in (Halteren and Teufel, 2003); as they do with factoids, we find increasing numbers of SCUs as the pool of summaries grows."*

P13-1020 → P11-1049 (Similar? Usage?): *"We use the same splits as previous work (Berg-Kirkpatrick et al., 2011; Woodsend and Lapata, 2012): the non-update portions of TAC-2009 for training and TAC-2008 for testing."* (Again a borderline case between similarity and usage classes)

F.5 Semantic Main Path Network: Delete Neutral and Future Work Citations

On the semantic main path network extracted by deleting neutral and future work citations (Figure 17 in the main text), we present some additional citation context excerpts of the annotated citation edges below.

D08-1081 → P08-1041 (Usage): *"For evaluating email thread summaries, we follow Carenini et al. (2008) by implementing their pyramid precision scheme, inspired by Nenkova's pyramid scheme (2004)."*

P10-1058 → W09-1801 (Similar): *"Our own work is closest to Martins and Smith (2009)."*

P10-1058 → D07-1047 (Comparison): *"Svore et al. (2007) were the first to foreground the highlight generation task which we adopt as an evaluation testbed for our model. Their approach is however a purely extractive one. Using* *an algorithm based on neural networks and third-party resources (e.g., news query logs and Wikipedia entries) they rank sentences and select the three highest scoring ones as story highlights. In contrast, we aim to generate rather than extract highlights."*

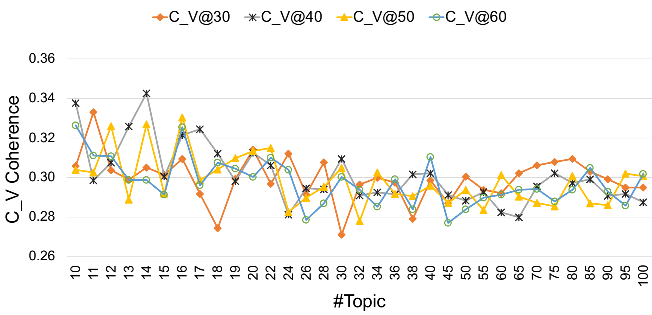
Below demonstrates the interesting difference between the paths from P11-1049 to P13-1020 in Figure 15 (by further adding similarity citations) and Figure 17 (by deleting neutral and future work citations). In Figure 15, the main path directly goes from P11-1049 to P13-1020 through a similarity citation, while in Figure 17, the path goes from P11-1049 through an intermediate paper D12-1022 to P11-1049.

P13-1020 → D12-1022 (Similar? Usage?): *"We use the same splits as previous work (Berg-Kirkpatrick et al., 2011; Woodsend and Lapata, 2012): the non-update portions of TAC-2009 for training and TAC-2008 for testing."* (Again a borderline case between similarity and usage classes)

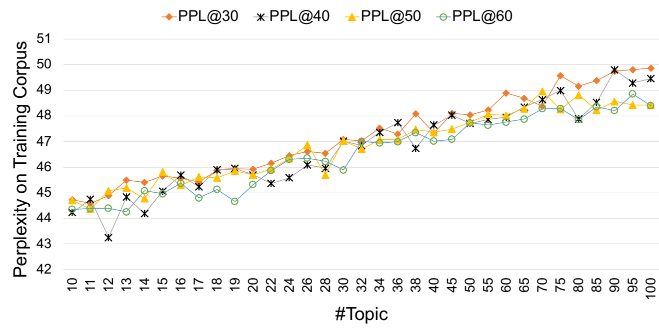
D12-1022 → P11-1049 (Comparison): *"Another closely related work also takes a max margin discriminative learning approach in the structural SVM framework (Berg-Kirkpatrick et al., 2011) ... However, they do not consider submodular functions, but instead ..."*

G. Supplementary Materials for Quantitative Analysis

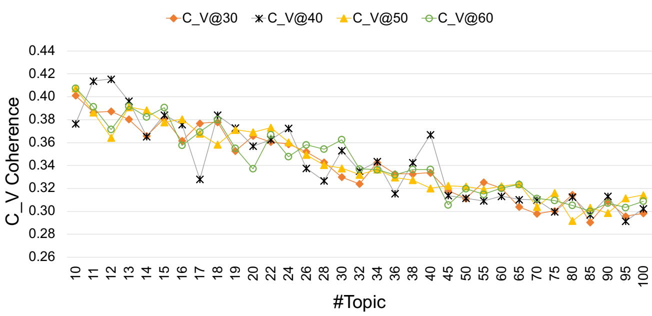
Two issues arose with topic modelling using Latent Dirichlet Allocation (LDA; Blei et al., 2003). First, we did not know the "right" value of *T* for each dataset, so for the purpose of getting robust results, we decided to evaluate using a range of reasonable *T* values and average. We set the minimum value of *T* to 10 according to the author's understanding of the studied subfields and the previous qualitative analysis results. Then, we trained LDA models based on different *T* values in {10, 11, …, 20, 22, …, 30, 35, 40, …, 100}. Second, to avoid overfitting LDA, we needed to chose a proper number of epochs for training. We tested different values of *P* in {30, 40, 50, 60}. To choose the "best" epoch, *C_V coherence* (Röder et al., 2015), also calculated using Gensim, was used to quantify the qualities of trained LDA models. For `AANPar` and `AANMT`, 26 and 20 were selected respectively as the upperbound of *T* because C_V coherence values dropped down to a lower plateau when *T* grew beyond the threshold value (see Supplementary Figure S1(c-d)). For **`AANSum`**, however, we further analysed the perplexity curve and selected 20 as the upperbound of *T* (see Supplementary Figure S1(a-b)). 40, 50, and 50 were decided as the "right" numbers of epochs for `AANSum`, `AANPar` and `AANMT` respectively. Supplementary Table S15 shows the C_V coherence averaged over all *T* values for each *P* value, based on which we inferred these values.



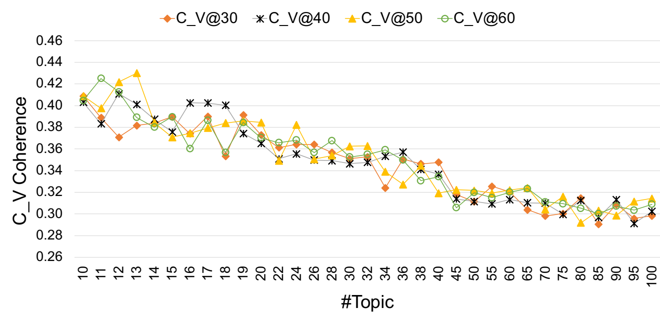
(a) CV coherence of LDA models on `AANSum`.



(b) Perplexity of LDA models on `AANSum`.



(c) CV coherence of LDA models on `AANPar`.



(b) CV coherence of LDA models on `AANMT`.

Figure S1. Selection of LDA-Based Topic Models for Quantitative MPA Evaluation.

Table S15. Selection of "Best" Number of Epoch for Quantitative MPA Evaluation.

	`AANSum`	`AANPar`	`AANMT`
C_V@P=30	0.3033	0.3720	0.3823
C_V@P=40	**0.3164**	0.3744	0.3915
C_V@P=50	0.3087	**0.3747**	**0.3928**
C_V@P=60	0.3064	0.3720	0.3872
Number of Topics	{10, 11, …, 20}	{10, 11, 20, 22, …, 26}	{10, 11, …, 20}
"Best" Epoch	40	50	50

Table S18. Topical Coverage of Main Path Networks on AANMT (T = 20, P = 50).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	cov_{tpk}	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%
10	0.0528	0.1073	103.22%	0.0843	59.66%	0.0773	46.40%	0.0860	62.88%	0.0722	36.74%	0.0765	44.89%
11	0.1361	0.0867	-36.30%	0.0541	-60.25%	0.0793	-41.73%	0.0594	-56.36%	0.0948	-30.35%	0.0390	-71.34%
12	0.0576	0.1134	96.88%	0.0627	8.85%	0.0691	19.97%	0.0709	23.09%	0.0727	26.22%	0.0529	-8.16%
13	0.0910	0.0903	-0.77%	0.0748	-17.80%	0.0823	-9.56%	0.0830	-8.79%	0.0478	-47.47%	0.0613	-32.64%
14	0.0488	0.0734	50.41%	0.0636	30.33%	0.0669	37.09%	0.0650	33.20%	0.0595	21.93%	0.0549	12.50%
15	0.0510	0.0591	15.88%	0.0680	33.33%	0.0853	67.25%	0.0571	11.96%	0.0565	10.78%	0.0441	-13.53%
16	0.0792	0.0758	-4.29%	0.0560	-29.29%	0.0575	-27.40%	0.0521	-34.22%	0.0403	-49.12%	0.0353	-55.43%
17	0.0419	0.0547	30.55%	0.0369	-11.93%	0.0494	17.90%	0.0391	-6.68%	0.0576	37.47%	0.0410	-2.15%
18	0.0566	0.0693	22.44%	0.0575	1.59%	0.0590	4.24%	0.0539	-4.77%	0.0327	-42.23%	0.0352	-37.81%
19	0.0826	0.0589	-28.69%	0.0662	-19.85%	0.0657	-20.46%	0.0530	-35.84%	0.0855	3.51%	0.0531	-35.71%
20	0.0683	0.0842	23.28%	0.0544	-20.35%	0.0745	9.08%	0.0633	-7.32%	0.0610	-10.69%	0.0529	-22.55%
Mean			24.78%		-2.34% ↓		9.34% ↑		-2.08% ↓		-3.93%		-20.18% ↓

Table S16. Main Path Network Topical Coverage Evaluation Results on AANSum (T = 20, P = 40).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	cov_{tpk}	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%
10	0.0653	0.0583	-10.72%	0.0626	-4.13%	0.0912	39.66%	0.0628	-3.83%	0.0395	-39.51%	0.0456	-30.17%
11	0.0705	0.0787	11.63%	0.0738	4.68%	0.0974	38.16%	0.0614	-12.91%	0.0831	17.87%	0.0595	-15.60%
12	0.0562	0.0630	12.10%	0.0703	25.09%	0.0413	-26.51%	0.0458	-18.51%	0.0649	15.48%	0.0451	-19.75%
13	0.0822	0.0904	9.98%	0.0531	-35.40%	0.0569	-30.78%	0.0519	-36.86%	0.0744	-9.49%	0.0449	-45.38%
14	0.0564	0.0667	18.26%	0.0889	57.62%	0.0839	48.76%	0.0662	17.38%	0.0734	30.14%	0.0496	-12.06%
15	0.0686	0.0781	13.85%	0.0554	-19.24%	0.0704	2.62%	0.0544	-20.70%	0.0698	1.75%	0.0383	-44.17%
16	0.0476	0.0511	7.35%	0.0532	11.76%	0.0453	-4.83%	0.0357	-25.00%	0.0570	19.75%	0.0303	-36.34%
17	0.0427	0.0641	50.12%	0.0316	-26.00%	0.0532	24.59%	0.0326	-23.65%	0.0561	31.38%	0.0330	-22.72%
18	0.0518	0.0507	-2.12%	0.0650	25.48%	0.0895	72.78%	0.0547	5.60%	0.0650	25.48%	0.0479	-7.53%
19	0.0482	0.0353	-26.76%	0.0402	-16.60%	0.0542	12.45%	0.0424	-12.03%	0.0380	-21.16%	0.0369	-23.44%
20	0.0829	0.0754	-9.05%	0.0557	-32.81%	0.0632	-23.76%	0.0525	-36.67%	0.0713	-13.99%	0.0542	-34.62%
Mean			6.79%		-0.87% ↓		13.92% ↑		-15.20% ↓		5.25%		-26.53% ↓

Table S17. Topical Coverage of Main Path Networks on AANPar (T = 26, P = 50).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	cov_{tpk}	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%	cov_{tpk}	Δ%
10	0.0506	0.0515	1.78%	0.0264	-47.83%	0.0407	-19.57%	0.0301	-40.51%	0.0687	35.77%	0.0266	-47.43%
11	0.0377	0.0324	-14.06%	0.0550	45.89%	0.0445	18.04%	0.0296	-21.49%	0.0518	37.40%	0.0246	-34.75%
12	0.0521	0.0870	66.99%	0.0273	-47.60%	0.0323	-38.00%	0.0381	-26.87%	0.0778	49.33%	0.0400	-23.22%
13	0.0655	0.0816	24.58%	0.0471	-28.09%	0.0452	-30.99%	0.0435	-33.59%	0.0989	50.99%	0.0490	-25.19%
14	0.0636	0.0766	20.44%	0.0505	-20.60%	0.0357	-43.87%	0.0414	-34.91%	0.0587	-7.70%	0.0388	-38.99%
15	0.0480	0.1023	113.13%	0.0507	5.63%	0.0366	-23.75%	0.0441	-8.13%	0.0687	43.13%	0.0377	-21.46%
16	0.0364	0.0473	29.95%	0.0515	41.48%	0.0394	8.24%	0.0421	15.66%	0.0424	16.48%	0.0276	-24.18%
17	0.0694	0.0717	3.31%	0.0260	-62.54%	0.0424	-38.90%	0.0321	-53.75%	0.0577	-16.86%	0.0313	-54.90%
18	0.0470	0.0650	38.30%	0.0632	34.47%	0.0478	1.70%	0.0320	-31.91%	0.0726	54.47%	0.0342	-27.23%
19	0.0454	0.0504	11.01%	0.0470	3.52%	0.0371	-18.28%	0.0361	-20.48%	0.0573	26.21%	0.0385	-15.20%
20	0.0638	0.0877	37.46%	0.0476	-25.39%	0.0424	-33.54%	0.0329	-48.43%	0.0796	24.76%	0.0349	-45.30%
22	0.0916	0.0595	-35.04%	0.0536	-41.48%	0.0528	-42.36%	0.0450	-50.87%	0.0948	3.49%	0.0489	-46.62%
24	0.0605	0.0720	19.01%	0.0554	-8.43%	0.0383	-36.69%	0.0423	-30.08%	0.0760	25.62%	0.0422	-30.25%
26	0.0485	0.0554	14.23%	0.0529	9.07%	0.0440	-9.28%	0.0343	-29.28%	0.0566	16.70%	0.0352	-27.42%
Mean			23.65%		-10.14% ↓		-21.95% ↓		-29.62% ↓		*25.70%*		-33.01% ↓

Table S19. Topical Coherence of Main Path Networks on AANSum ($T = 20$, $P = 40$).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	coh_{tpk}	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%
10	0.5682	0.5227	−8.01%	0.5731	+0.86%	0.5584	−1.72%	0.5571	−1.95%	0.5610	−1.27%	0.5623	−1.04%
11	0.5061	0.5229	+3.32%	0.4932	−2.55%	0.4791	−5.33%	0.4809	−4.98%	0.5021	−0.79%	0.4922	−2.75%
12	0.4236	0.3817	−9.89%	0.3934	−7.13%	0.4079	−3.71%	0.3985	−5.93%	0.4151	−2.01%	0.4091	−3.42%
13	0.5912	0.5888	−0.41%	0.5967	+0.93%	0.6067	+2.62%	0.5875	−0.63%	0.5879	−0.56%	0.5909	−0.05%
14	0.4893	0.4483	−8.38%	0.4635	−5.27%	0.4655	−4.86%	0.4718	−3.58%	0.4707	−3.80%	0.4785	−2.21%
15	0.5298	0.5442	+2.72%	0.5394	+1.81%	0.5424	+2.38%	0.5430	+2.49%	0.5682	+7.25%	0.5555	+4.85%
16	0.5767	0.5905	+2.39%	0.5893	+2.18%	0.6000	+4.04%	0.5883	+2.01%	0.5913	+2.53%	0.5914	+2.55%
17	0.5593	0.5208	−6.88%	0.5499	−1.68%	0.5438	−2.77%	0.5452	−2.52%	0.5634	+0.73%	0.5539	−0.97%
18	0.5941	0.6160	+3.69%	0.6307	+6.16%	0.6283	+5.76%	0.6333	+6.60%	0.5849	−1.55%	0.6199	+4.34%
19	0.5913	0.5360	−9.35%	0.5639	−4.63%	0.5423	−8.29%	0.5492	−7.12%	0.5809	−1.76%	0.5603	−5.24%
20	0.6403	0.6134	−4.20%	0.6086	−4.95%	0.5963	−6.87%	0.6104	−4.67%	0.6302	−1.58%	0.6181	−3.47%
Mean	0.5518	0.5350	−3.18%	0.5456	−1.30%↑	0.5428	−1.70% ↓	0.5423	−1.84%↓	0.5505	−0.26%	0.5484	−0.67% ↓

Table S20. Topical Coherence of Main Path Networks on AANPar ($T = 26$, $P = 50$).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	coh_{tpk}	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%
10	0.3853	0.3587	−6.90%	0.3964	+2.88%	0.3894	+1.06%	0.3813	−1.04%	0.3830	−0.60%	0.3825	−0.73%
11	0.3917	0.3909	−0.20%	0.4249	+8.48%	0.3981	+1.63%	0.3973	+1.43%	0.3969	+1.33%	0.3955	−0.97%
12	0.4211	0.4230	+0.45%	0.4236	+0.59%	0.4219	+0.19%	0.4244	+0.78%	0.4303	+2.18%	0.4262	+1.21%
13	0.3629	0.3414	−5.92%	0.3574	−1.52%	0.3445	−5.07%	0.3439	−5.24%	0.3442	−5.15%	0.3447	−5.02%
14	0.3963	0.3820	−3.61%	0.4007	+1.11%	0.3918	−1.14%	0.3892	−1.79%	0.3859	−2.62%	0.3883	−2.02%
15	0.4547	0.4232	−6.93%	0.4514	−0.73%	0.4539	−0.18%	0.4413	−2.95%	0.4446	−2.22%	0.4416	−2.88%
16	0.4610	0.4556	−1.17%	0.4679	+1.50%	0.4541	−1.50%	0.4542	−1.48%	0.4547	−1.37%	0.4552	−1.26%
17	0.4767	0.4633	−2.81%	0.4806	+0.82%	0.4746	−0.44%	0.4695	−1.51%	0.4825	+1.22%	0.4726	−0.86%
18	0.4180	0.4372	+4.59%	0.4298	+2.82%	0.4220	+0.96%	0.4288	+2.58%	0.4268	+2.11%	0.4279	+2.37%
19	0.4471	0.4730	+5.79%	0.4710	+5.35%	0.4597	+2.82%	0.4629	+3.53%	0.4463	−0.18%	0.4575	+2.33%
20	0.5330	0.5273	−1.07%	0.5553	+4.18%	0.5355	+0.47%	0.5340	+0.19%	0.5221	−2.05%	0.5307	−0.43%
22	0.4921	0.5074	+3.11%	0.5045	+2.52%	0.5043	+2.48%	0.5037	+2.36%	0.4975	+1.10%	0.5021	+2.03%
24	0.5361	0.5034	−6.10%	0.5346	−0.28%	0.5274	−1.62%	0.5188	−3.23%	0.5346	−0.28%	0.5241	−2.24%
26	0.5302	0.5411	+2.06%	0.5417	+2.17%	0.5283	−0.36%	0.5344	+0.79%	0.5119	−3.45%	0.5291	−0.21%
Mean	0.4504	0.4448	−1.34%	0.4600	+2.14% ↑	0.4504	−0.05% ↓	0.4488	−0.40% ↓	0.4772	−0.71%	0.4484	−0.48% ↑

Table S21. Topical Coherence of Main Path Networks on AANMT ($T = 20$, $P = 50$).

T	MPN	add_Ext_Mot		plus_add_Use		plus_add_Sim		add_Combined		del_Bkg_Fut		del_Combined	
	coh_{tpk}	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%	coh_{tpk}	Δ%
10	0.3516	0.3267	−7.08%	0.3326	−5.40%	0.3267	−11.35%	0.3287	−6.51%	0.3347	−4.81%	0.3309	−5.89%
11	0.3650	0.3906	+7.01%	0.3973	+8.85%	0.3906	−0.25%	0.3866	+5.92%	0.3695	+1.23%	0.3806	+4.27%
12	0.3757	0.3670	−2.32%	0.3794	+0.98%	0.3670	−6.65%	0.3611	−3.89%	0.3732	−0.67%	0.3653	−2.77%
13	0.4174	0.4119	−1.32%	0.4180	+0.14%	0.4119	−5.94%	0.4045	−3.09%	0.4064	−2.64%	0.4042	−3.16%
14	0.4450	0.4283	−3.75%	0.4455	+0.11%	0.4283	−6.56%	0.4262	−4.22%	0.4424	−0.58%	0.4330	−2.70%
15	0.4256	0.4254	−0.05%	0.4531	+6.46%	0.4254	+1.17%	0.4320	+1.50%	0.4277	+0.49%	0.4299	+1.01%
16	0.4230	0.4125	−2.48%	0.4247	+0.40%	0.4125	−3.78%	0.4163	−1.58%	0.4218	−0.28%	0.4186	−1.04%
17	0.5051	0.4855	−3.88%	0.5002	−0.97%	0.4855	−4.71%	0.4880	−3.39%	0.5059	+0.16%	0.4936	−2.28%
18	0.4935	0.4804	−2.65%	0.4962	+0.55%	0.4804	−5.11%	0.4781	−3.12%	0.4948	+0.26%	0.4852	−1.68%
19	0.4516	0.4911	+8.75%	0.4957	+9.77%	0.4911	+0.82%	0.4750	+5.18%	0.4606	+1.99%	0.4693	+3.92%
20	0.5064	0.4673	−7.72%	0.4903	−3.18%	0.4673	−6.38%	0.4746	−6.28%	0.4919	−2.86%	0.4820	−4.82%
Mean	0.4327	0.4261	−1.41%	0.4394	+1.61% ↑	0.4138	−4.43% ↓	0.4246	−1.77% ↓	0.4299	−0.70%	0.4266	−1.38% ↑

H. MORE REFERENCES

Blei, D., Ng., A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022. https://www.jmlr.org/papers/v3/blei03a.html

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. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL'22)*, 1875–1888. https://aclanthology.org/2022.naacl-main.137/

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

Röder, M., Both, A., & Hineburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM'15)*, 399–408. https://doi.org/10.1145/2684822.2685324

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