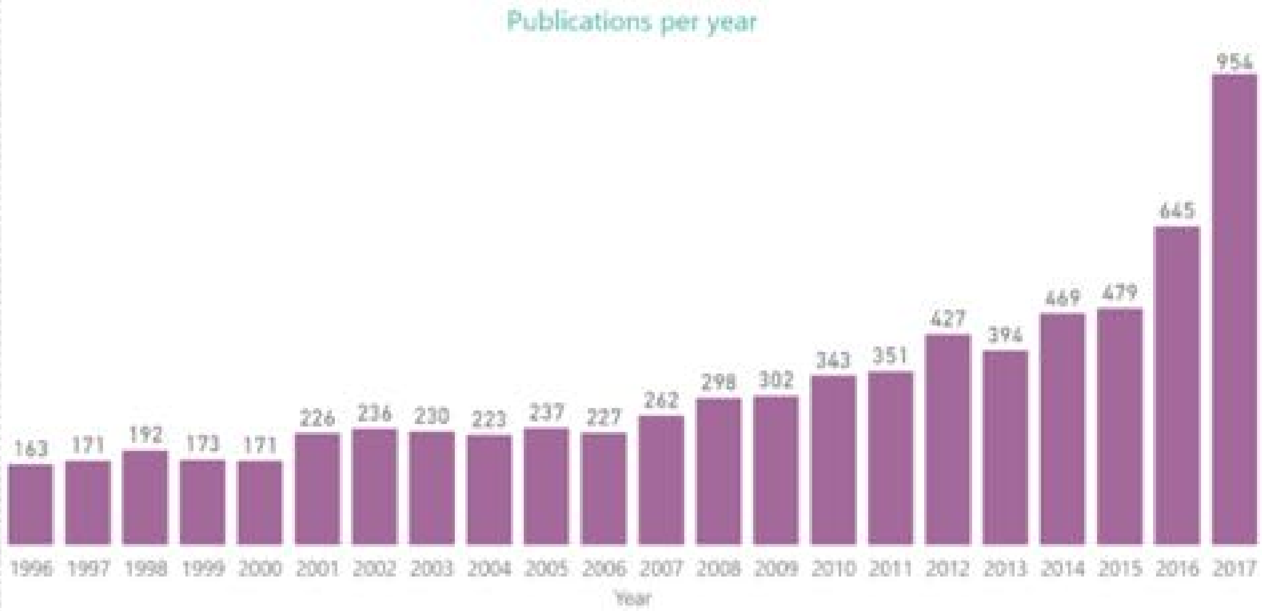
A Network Tour of NIPS conference Papers

Yawen Hou
Peilin Kang
Yueran Liang
Zhechen Su

2020.01

DATASET: 1987- 2018



- Neural Information Processing Systems (NIPS) conference
- Topics:
  - Machine learning
    - Reinforcement learning
    - Deep learning
    - etc...
  - Cognitive science
  - Information theory
  - etc...
- **11630** authors who published **7241** documents in total.
- Scrapped using Python script[1]
  - **Authors**.csv: authors_id, Name of authors
  - **Papers_authors**.csv: author_id, paper_id
  - **Papers**.csv: paper_id, [strings] (paper)

[1]https://www.kaggle.com/benhamner/nips-papers

Preprocessing

Bag of Words

Stop words

Bi-grams

Stemming

Transform each sentence into a vector of words.

- Remove punctuations & space

Nobody knows how ancient people started using fire.

↓

{Nobody, knows, how, ancient, people, started, using, fire}

Preprocessing

Bag of Words

Stop words

Bi-grams

Stemming

Remove words appearing with high frequency.

articles, prepositions, adverbs, conjunctions, (verbs)

{Nobody, ***knows***, ***how***, ancient, people, started, ***using***, fire}

↓

{Nobody, ancient, people, started, fire}.

Preprocessing

Bag of Words

Stop words

Bi-grams

Stemming

Treat each meaningful pair of words as one.

- Link pair with an underscore "_"
- Filter with low_threshold and high_threshold
- New York ⇒ New_York
- ~~et al.~~

{Nobody, ancient, people, started, fire}

↓

{Nobody, **ancient_people**, started, fire}

Preprocessing

Bag of Words

Stop words

Bi-grams

Stemming

Reduce inflected words to their stem.

Remove variation in plurality, verb tense, noun, adjective, capitalization

done = doing, women = woman

{Nobody, ancient_people, started, fire}

↓

{nobody, ancient_people, **start**, fire}

LDA

Latent Dirichlet allocation

- Discovering hidden topical patterns that are present across the collection
- Annotating documents according to these topics
- Using these annotations to organize, search and summarize texts

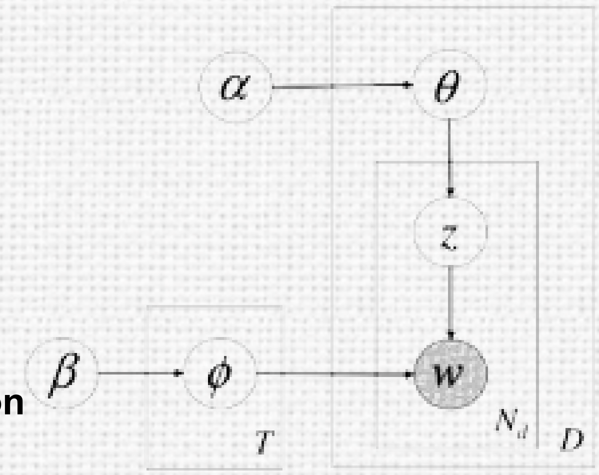
Docs + Topic_num

↓

LDA

↓

- **Topics representation**
  (Normed Vector of words)
- **Documents representation**
  (Normed Vector of topics)



David Blei, Andrew Ng, and Michael Jordan. Latent dirichlet allocation. Journal of Machine Learning Research, 3:993–1022, 05 2003.

ATM

Author Topic Model

- Extra author preference
- A document = a mixture of the topic distributions affiliated with the authors.
- author to documents is 1 to n
  author to topics is 1 to n

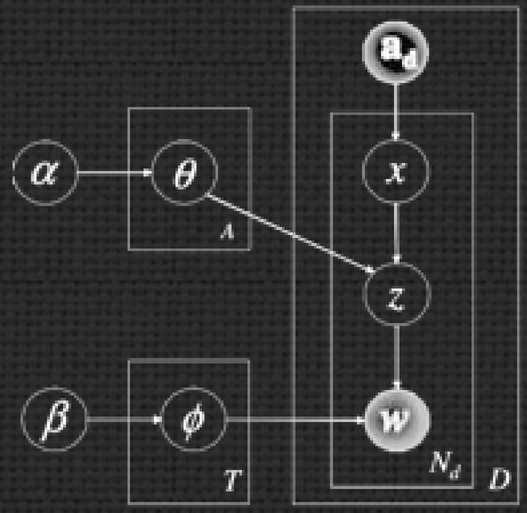
Docs + Topic_num

↓

ATM

↓

- **Topics representation**
  (Normed Vector of words)
- **Author representation**
  (Normed Vector of topics)



Michal Rosen-Zvi et.al. The author-topic model for authors and documents. Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence,

LDA

Latent Dirichlet allocation

- Discovering hidden topical patterns that are present across the collection
- Annotating documents according to these topics
- Using these annotations to organize, search and summarize texts

ATM

Author Topic Model

- Extra author preference
- A document = a mixture of the topic distributions affiliated with the authors.
- author to documents is 1 to n

author to topics is 1 to n

Topic 1: Food

0.4*peanut, 0.3*almonds…

Topic 2: Pet

0.45*cat, 0.45*dog…

Author A:

0.9*food, 0.1*pet…

Author B:

0.9*pet, 0.1*food

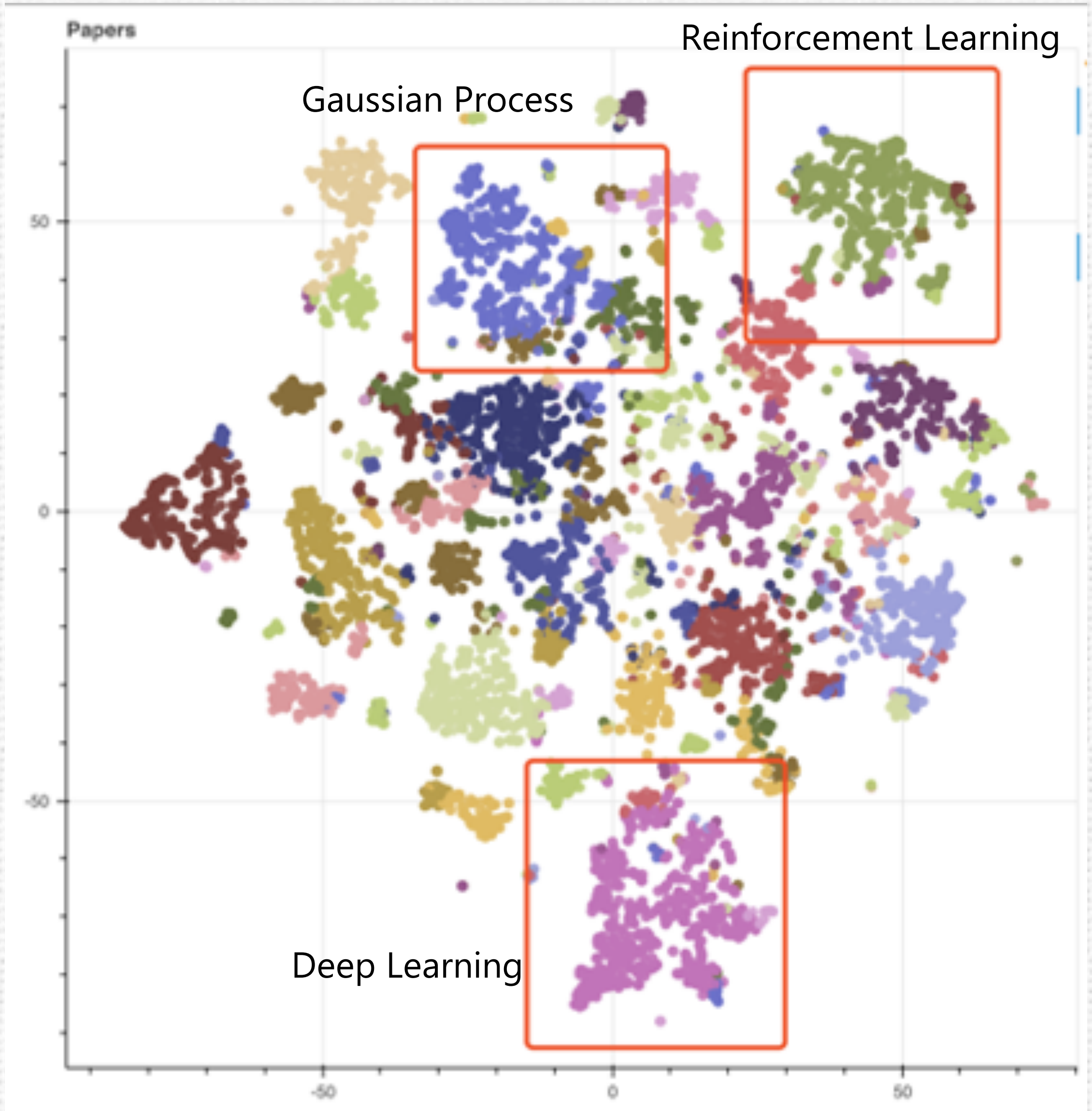
LDA Topic Extraction

- Derive the topics using the LDA models
- Each topic a named label for convenient description.
- List the top-20 most frequent topics

Topic Label	High-frequency Terms									
Supervised Learning	label	classifier	margin	active	hypothesis	target	classifi	multi	boost	binary
Optimization	gradient	convex	objective	stochastic	dual	min	smooth	minimization	norm	converge
Gaussian Processes	posterior	bayesian	inference	likelihood	mixture	latent	variational	conditional	probabilistic	
Signal Processing	signal	filter	source	frequency	brain	channel	coefficient	correlation	speech	time_serie
Information Theory	estimator	density	finite	entropy	bias	continuous	asymptotic	theoretical	converge	uniform
Reinforcement Learning	policy	action	reward	agent	game	reinforcement_learn	player	environment	transition	strategy
Deep Learning	layer	deep	architecture	neural_network	convolutional	preprint	cnn	recurrent	gradient	rnn
Deep Learning	rule	neural$_network$	net	layer	hide	architecture	generalization	dynamic	activation	hidden_unit
Graph theory	graph	node	tree	edge	vertex	inference	partition	path	product	degree
Matrix and Tensor Factorization	sparse	column	rank	norm	sparsity	row	tensor	entry	recover	low_rank
Unsupervised Learning	cluster	distance	metric	similarity	manifold	embed	clustering	projection	nearest_neighbor	eigenvector
Computer vision	object	pixel	segmentation	recognition	face	detection	visual	scene	region	shape
Bandit algorithms	bind	regret	lemma	online	lower_bound	round	arm	upper_bound	bandit	strategy
Kernel methods	kernel	regression	svm	regularization	selection	operator	nonlinear	support_vector	polynomial	functional
Human Learning	human	response	trial	target	subject	stimulus	visual	behavior	cue	account
Collaborative filtering	user	rank	item	group	query	score	event	individual	link	preference
Information Retrieval	word	topic	document	language	text	sentence	lda	length	character	semantic
Navigation and Planning	motion	position	location	region	trajectory	direction	field	move	surface	center
Neuroscience	neuron	cell	splike	activity	response	stimulus	circuit	synaptic	dynamic	population
Distributed Computing	memory	search	bit	distribute	code	parallel	block	communication	scheme	store

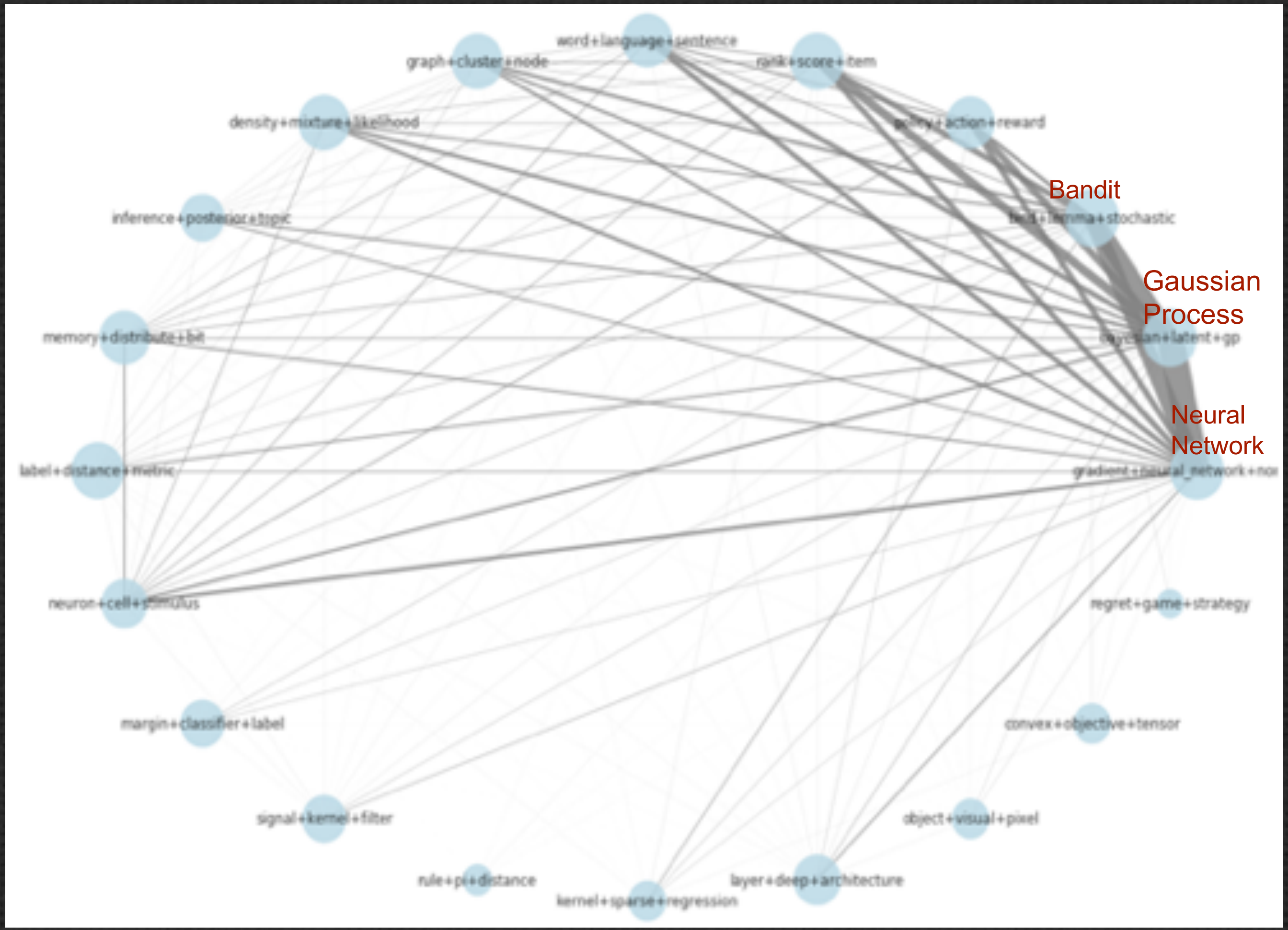
LDA Topic Distribution

- Apply t-SNE for the exploitation of the distribution of topics of papers over years
- 3 largest clusters of topics are Deep Learning, Gaussian Process and Reinforcement Learning.



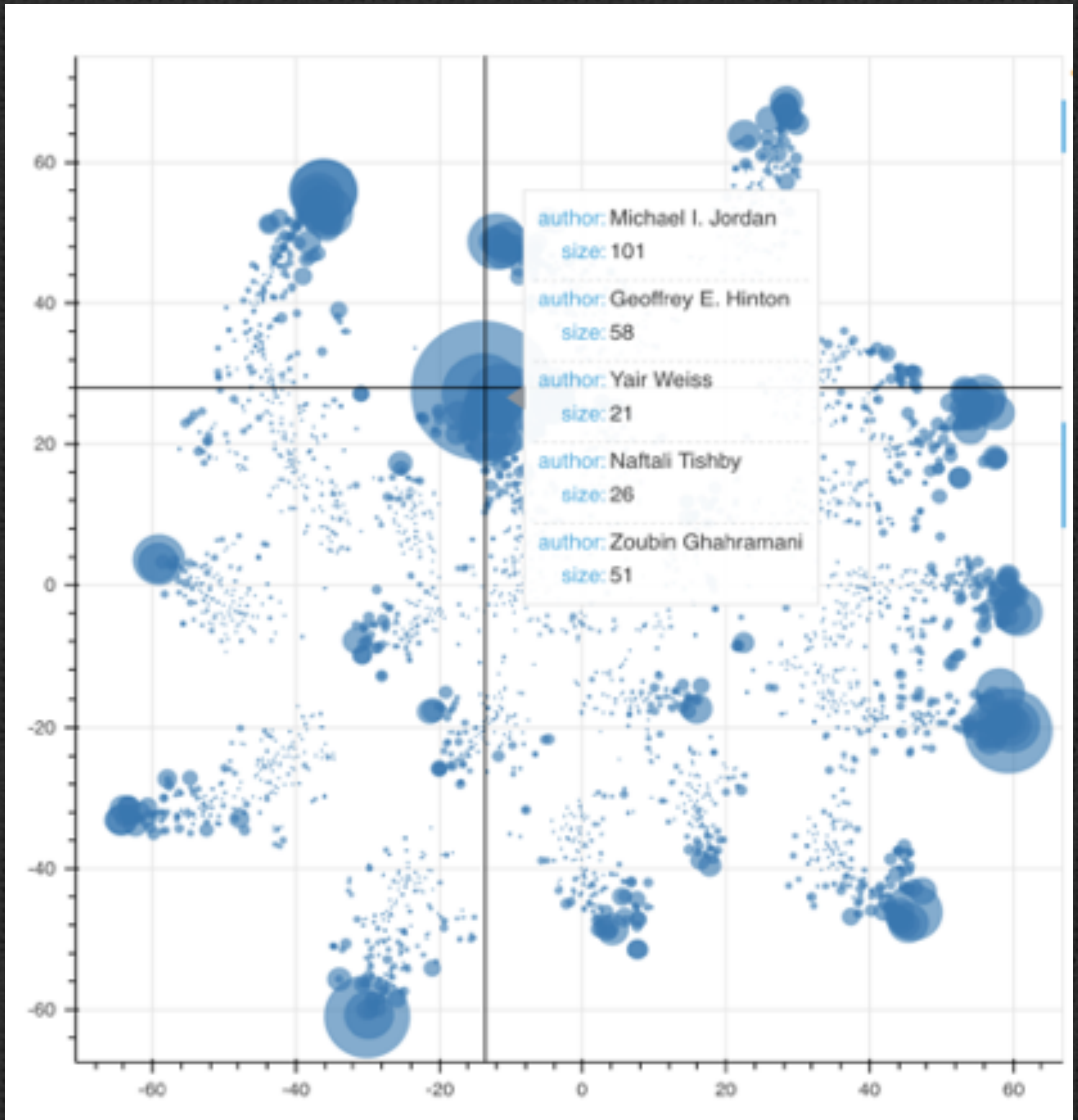
ATM Topic Network

- Derive the topics and the topic preference of each author using the ATM model.
- Discover domain topic of each researcher.
- Each node represent a topic (a mixture of words).
- Connect two nodes if two authors working on these two topics have collaborated together in the writing of a paper.
- Three most frequently collaborated topics (Bandit, Gaussian Process, Neural Network)



Dimensionality Reduction (t-SNE) on topic preference

- The position of the center of each circle represents the preference of each author.
- The size of the circle represents the number of the papers this author has written.
- The larger the circle, the more papers they write.
- If the centers of two circles are very close to each other, it means that the topic preferences of these two authors are very similar.



Topic Similarity

- Use the Hellinger distance to measuring the distance between two authors' topic distribution (i.e their topic preference)
- The higher the Hellinger distance, the more dissmilar two distributions are.

$$H(p,q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{K}(\sqrt{p_i} - \sqrt{q_i})^2}$$

- Convert into similarity

$$S(p,q) = \frac{1}{1 + H(p,q)}$$

- The score column indicates the similarity S(p=Terrence J. Sejnowski, q), where q represents the other authors. The higher the score, the more similar they are.
- The topic column indicates the topic preference of this author.
- They are all more or less interested to the topic 11.

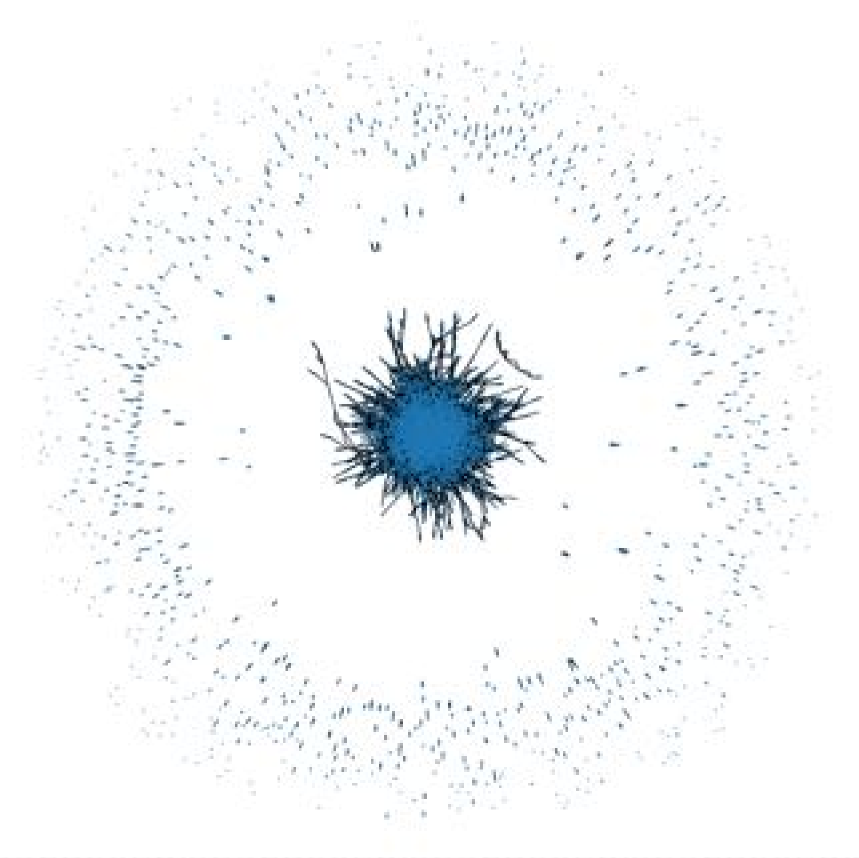
	Author	Score	Size	topic
2845	Terrence J. Sejnowski	1.000000	48	[(11, 0.9999290468440801)]
2258	Peter Dayan	0.999995	47	[(11, 0.9999435699752137)]
534	Christof Koch	0.999985	35	[(11, 0.9998864424808276)]
2680	Si Wu	0.999912	9	[(11, 0.9996814870352174)]
66	Alan F. Murray	0.999907	11	[(11, 0.999664936787143)]
251	Anthony M. Zador	0.999897	8	[(11, 0.999638815817049)]
523	Christian K. Machens	0.999875	7	[(11, 0.9995754359247159)]
2469	Rodney J. Douglas	0.999872	8	[(11, 0.9995657793713213)]
509	Chris Diorio	0.999842	7	[(11, 0.9994811398151168)]
1620	Kevin A. Archie	0.999781	3	[(11, 0.9993097480936701)]

Co-authorship Network

undirected weighted graph

	Description	Amount
nodes	each author	11630
edges	two authors have written the same paper	27935
weight of the edge	the number of papers written together	

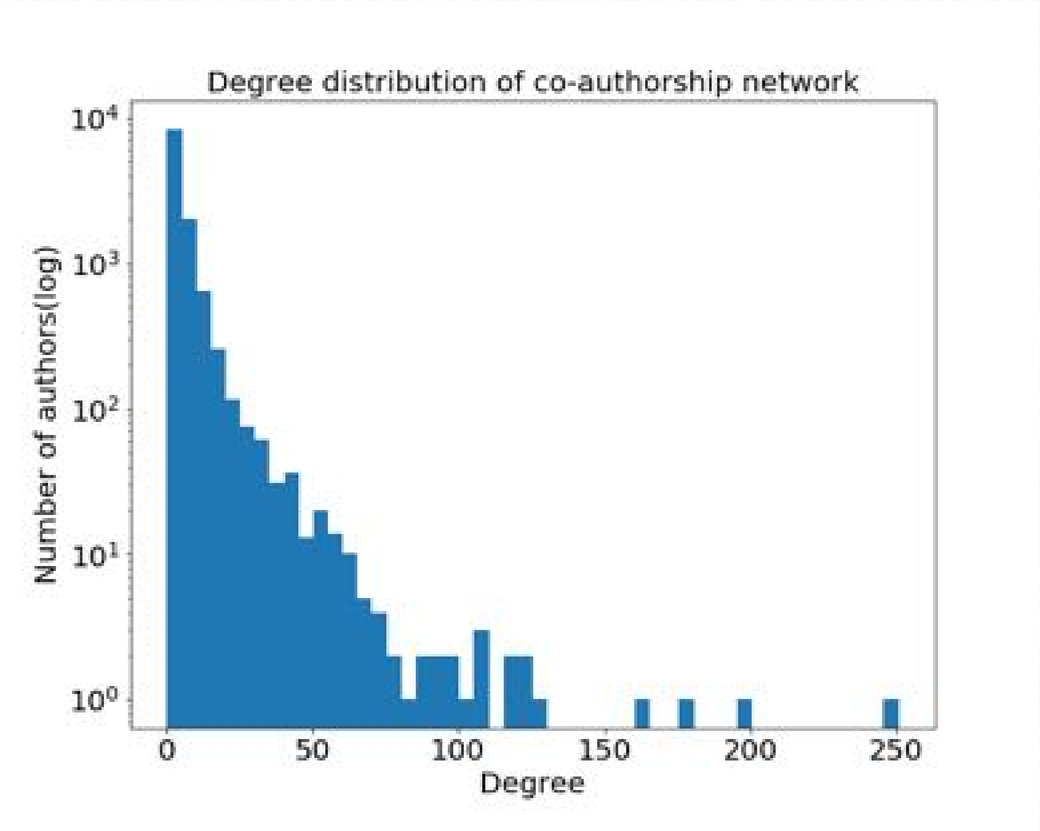
Co-authorship Network Exploration



- Average Degree: 4.8
- Sparsity: 0.000413
- Average clustering: 0.68
- Total number of components: 915
  The largest component contains: 77.2% authors

component_size	8978	21	20	18	17	16	14	13	12	11	10	9	8	7	6	5	4	3	2	1
number of components	1	1	1	1	1	1	2	2	4	4	2	5	7	24	26	44	95	192	291	211

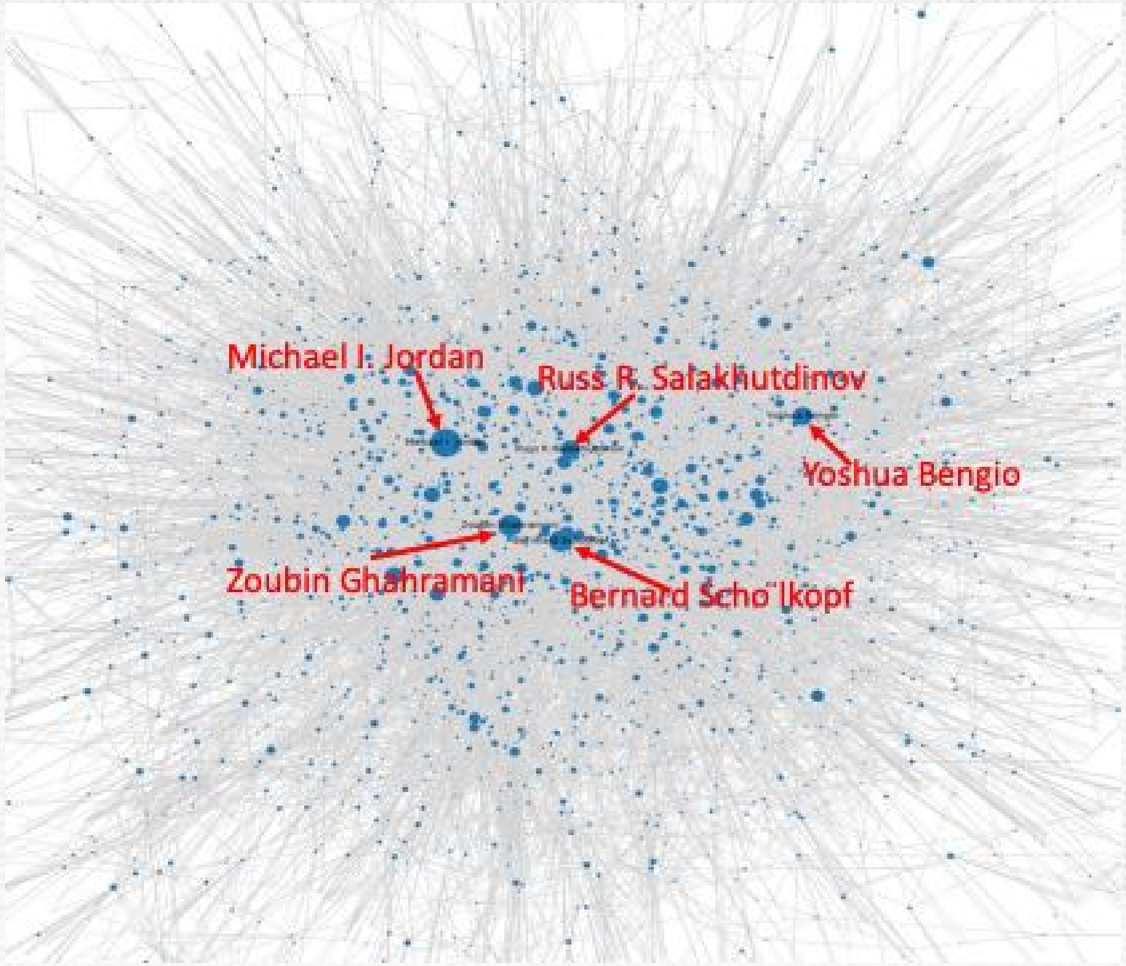
Co-authorship Network Exploration



- Power Law
- Most nodes in the co-authorship network have small degrees
- Only a few authors have large degrees
- Most authors have limited cooperation relationship between other authors

Largest connected component

Average shortest-path length: 5.95



Authors with Top-5 betweenness centrality

- Michael I. Jordan
- Bernard Scho¨lkopf
- Zoubin Ghahramani
- Yoshua Bengio
- Russ R. Salakhutdinov

Michael I. Jordan

FOLLOW

Professor of EECS and Professor of Statistics, University of California, Berkeley
Verified email at cs.berkeley.edu - Homepage

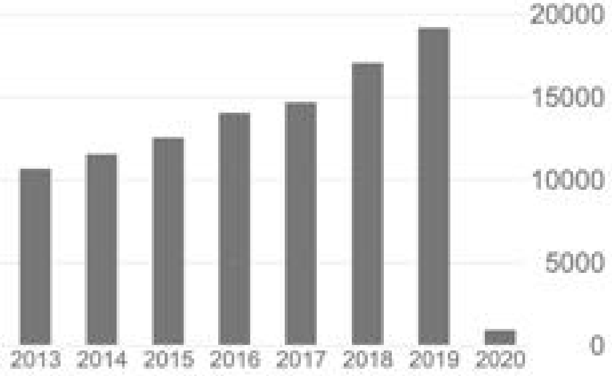
machine learning   statistics   computational biology   artificial intelligence   optimization

TITLE	CITED BY	YEAR
Latent dirichlet allocation DM Blei, AY Ng, MI Jordan Journal of machine Learning research 3 (Jan), 993-1022	30171	2003
On spectral clustering: Analysis and an algorithm AY Ng, MI Jordan, Y Weiss Advances in neural information processing systems, 849-856	8167	2002
Adaptive mixtures of local experts RA Jacobs, MI Jordan, SJ Nowlan, GE Hinton Neural computation 3 (1), 79-87	4126	1991
Sharing clusters among related groups: Hierarchical Dirichlet processes YW Teh, MI Jordan, MJ Beal, DM Blei Advances in neural information processing systems, 1385-1392	3969	2005

GET MY OWN PROFILE

Cited by   VIEW ALL

	All	Since 2015
Citations	170946	78539
h-index	161	109
i10-index	552	416



Conclusion

Topic Exploration

- LDA and ATM models
- Visualisation using t-SNE
- Inter-topic collaboration
- Author's topic preference
- Author recommendation based on topics

Co-authorship

- many authors have co-authored ⇒ large connected component
- a few hubs

THANKS

Questions?