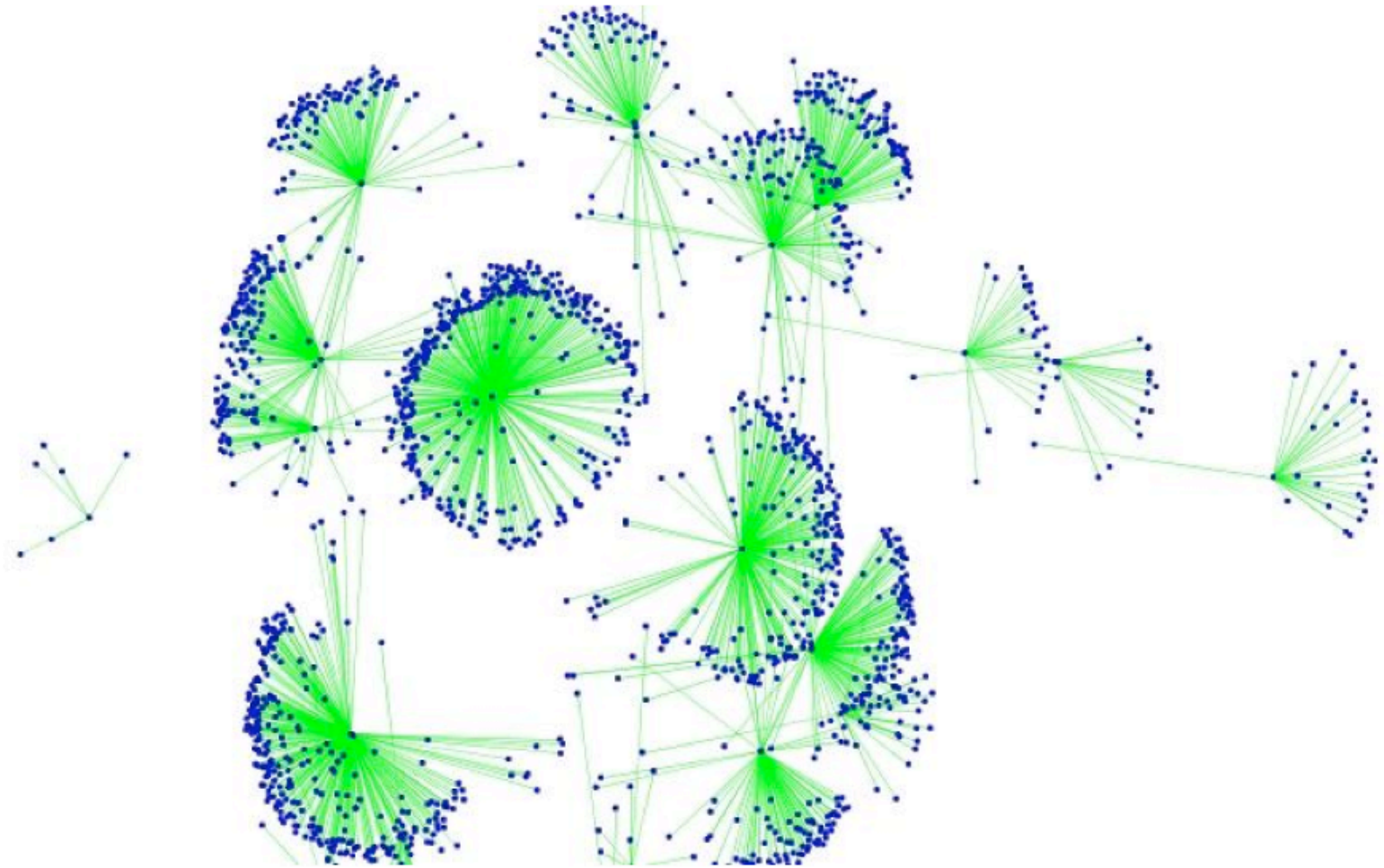
Predictive Social Network Analysis with Multi-Modal Data

Andrew McCallum

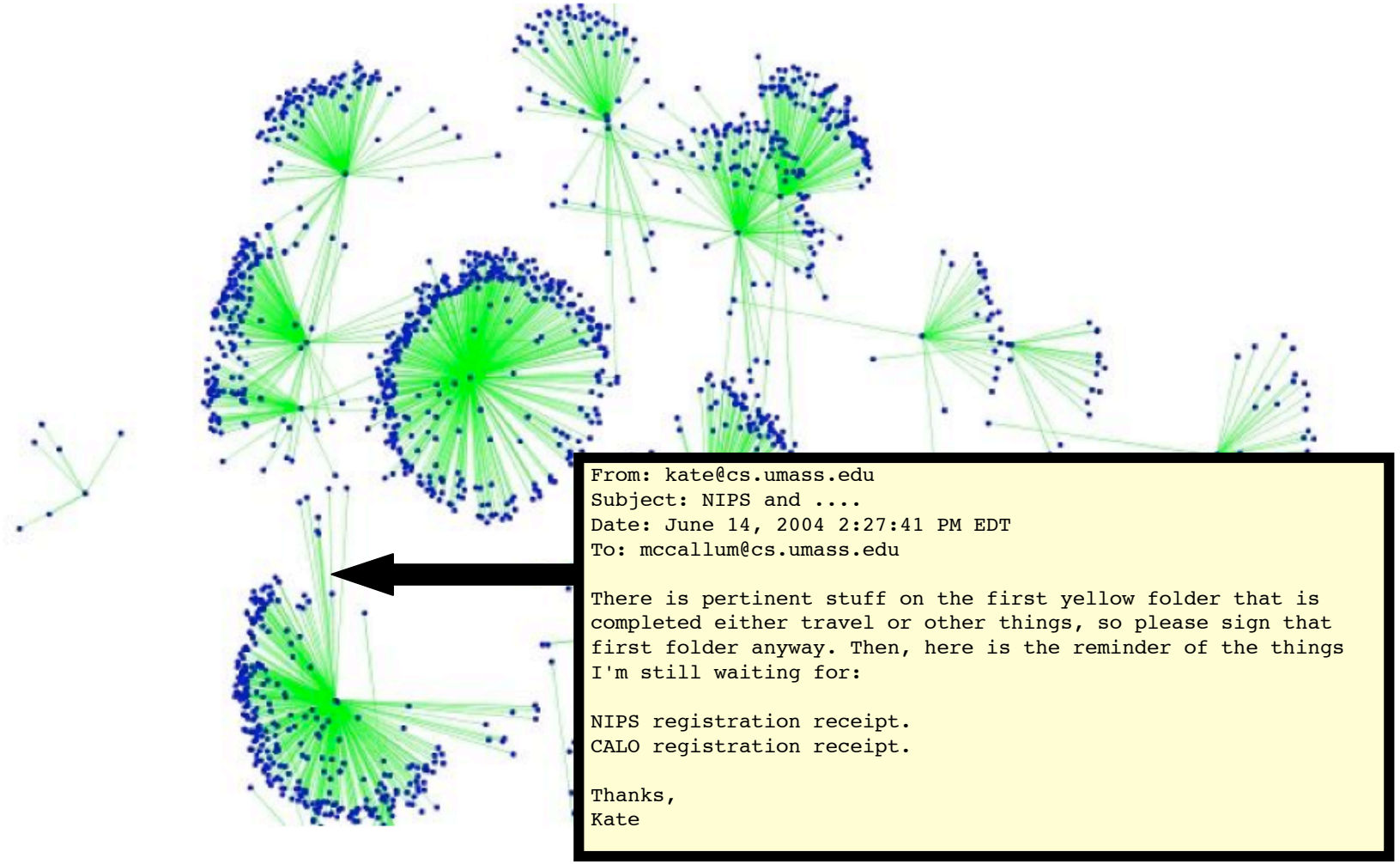
Computer Science Department
University of Massachusetts Amherst

Joint work with Wei Li, Xuerui Wang, Andres Corrada,
Chris Pal, Natasha Mohanty, David Mimno, Gideon Mann.

Social Network in an Email Dataset



Social Network in an Email Dataset



```
From: kate@cs.umass.edu
Subject: NIPS and ....
Date: June 14, 2004 2:27:41 PM EDT
To: mccallum@cs.umass.edu

There is pertinent stuff on the first yellow folder that is
completed either travel or other things, so please sign that
first folder anyway. Then, here is the reminder of the things
I'm still waiting for:

NIPS registration receipt.
CALO registration receipt.

Thanks,
Kate
```

Social Network Analysis
(Pattern Discovery in Networks)

Data:

Network Connectivity → **Network Connectivity + *Many Attributes***

Text, Timestamps, Authors,...

Objective:

Descriptive → **Predictive & Prescriptive**

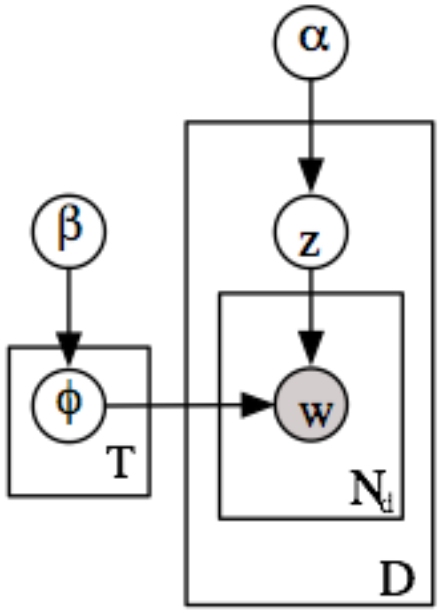
Small World, Betweeness Centrality,...

P(edge|...) P(attribute|...), P(group|...)
P(collaborator|...)

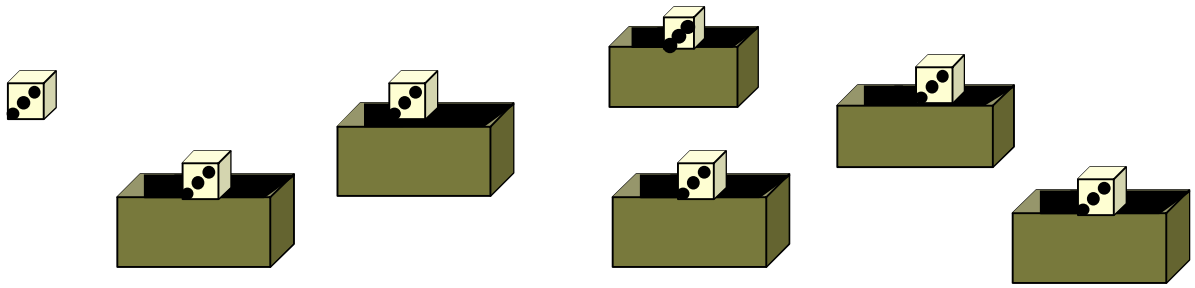
Methodology:

Direct Measures, Agglomerative & Spectral Clustering,... → **Generative, Latent Variable Models**

A Probabilistic Approach

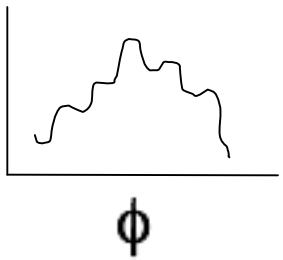


- Define a probabilistic generative model for documents.



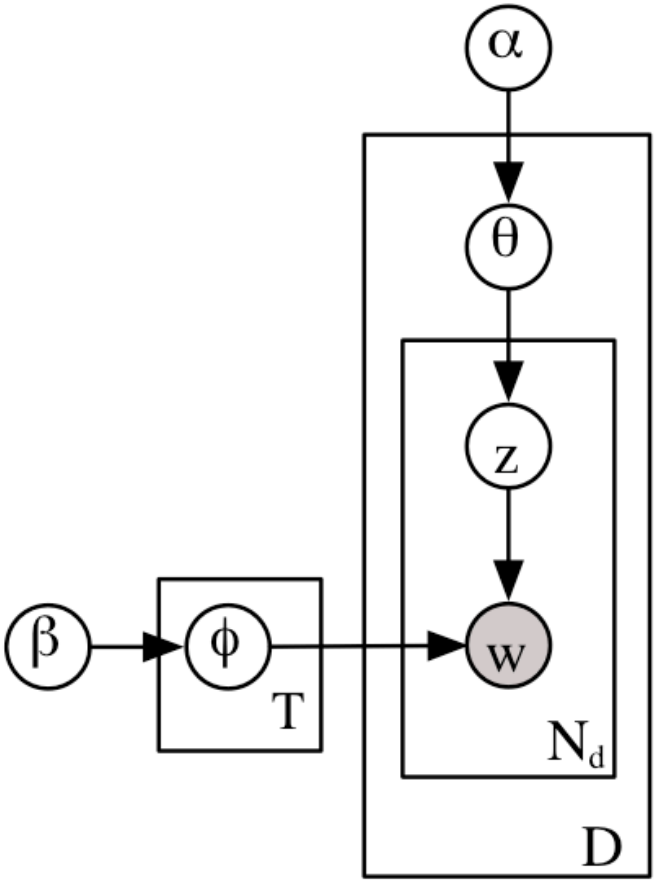
- Learn the parameters of this model by fitting them to the data and a prior.

$$\phi^* = \arg\max_\phi p(\phi|D_1D_2...) = p(D_1D_2...|\phi)\, p(\phi)$$



Clustering words into topics with Latent Dirichlet Allocation

[Blei, Ng, Jordan 2003]



Generative Process:

For each document:

Sample a distribution over topics, θ

For each word in doc

Sample a topic, *z*

Sample a word from the topic, *w*

Example:

70% Iraq war
30% US election

Iraq war

"bombing"

Example topics
induced from a large collection of text

DISEASE	WATER	MIND	STORY	FIELD	SCIENCE	BALL	JOB
BACTERIA	FISH	WORLD	STORIES	MAGNETIC	STUDY	GAME	WORK
DISEASES	SEA	DREAM	TELL	MAGNET	SCIENTISTS	TEAM	JOBS
GERMS	SWIM	DREAMS	CHARACTER	WIRE	SCIENTIFIC	FOOTBALL	CAREER
FEVER	SWIMMING	THOUGHT	CHARACTERS	NEEDLE	KNOWLEDGE	BASEBALL	EXPERIENCE
CAUSE	POOL	IMAGINATION	AUTHOR	CURRENT	WORK	PLAYERS	EMPLOYMENT
CAUSED	LIKE	MOMENT	READ	COIL	RESEARCH	PLAY	OPPORTUNITIES
SPREAD	SHELL	THOUGHTS	TOLD	POLES	CHEMISTRY	FIELD	WORKING
VIRUSES	SHARK	OWN	SETTING	IRON	TECHNOLOGY	PLAYER	TRAINING
INFECTION	TANK	REAL	TALES	COMPASS	MANY	BASKETBALL	SKILLS
VIRUS	SHELLS	LIFE	PLOT	LINES	MATHEMATICS	COACH	CAREERS
MICROORGANISMS	SHARKS	IMAGINE	TELLING	CORE	BIOLOGY	PLAYED	POSITIONS
PERSON	DIVING	SENSE	SHORT	ELECTRIC	FIELD	PLAYING	FIND
INFECTIOUS	DOLPHINS	CONSCIOUSNESS	FICTION	DIRECTION	PHYSICS	HIT	POSITION
COMMON	SWAM	STRANGE	ACTION	FORCE	LABORATORY	TENNIS	FIELD
CAUSING	LONG	FEELING	TRUE	MAGNETS	STUDIES	TEAMS	OCCUPATIONS
SMALLPOX	SEAL	WHOLE	EVENTS	BE	WORLD	GAMES	REQUIRE
BODY	DIVE	BEING	TELLS	MAGNETISM	SCIENTIST	SPORTS	OPPORTUNITY
INFECTIONS	DOLPHIN	MIGHT	TALE	POLE	STUDYING	BAT	EARN
CERTAIN	UNDERWATER	HOPE	NOVEL	INDUCED	SCIENCES	TERRY	ABLE

[Tennenbaum et al]

Example topics
induced from a large collection of text

DISEASE	WATER	MIND	STORY	**FIELD**	SCIENCE	BALL	JOB
BACTERIA	FISH	WORLD	STORIES	MAGNETIC	STUDY	GAME	WORK
DISEASES	SEA	DREAM	TELL	MAGNET	SCIENTISTS	TEAM	JOBS
GERMS	SWIM	DREAMS	CHARACTER	WIRE	SCIENTIFIC	FOOTBALL	CAREER
FEVER	SWIMMING	THOUGHT	CHARACTERS	NEEDLE	KNOWLEDGE	BASEBALL	EXPERIENCE
CAUSE	POOL	IMAGINATION	AUTHOR	CURRENT	WORK	PLAYERS	EMPLOYMENT
CAUSED	LIKE	MOMENT	READ	COIL	RESEARCH	PLAY	OPPORTUNITIES
SPREAD	SHELL	THOUGHTS	TOLD	POLES	CHEMISTRY	**FIELD**	WORKING
VIRUSES	SHARK	OWN	SETTING	IRON	TECHNOLOGY	PLAYER	TRAINING
INFECTION	TANK	REAL	TALES	COMPASS	MANY	BASKETBALL	SKILLS
VIRUS	SHELLS	LIFE	PLOT	LINES	MATHEMATICS	COACH	CAREERS
MICROORGANISMS	SHARKS	IMAGINE	TELLING	CORE	BIOLOGY	PLAYED	POSITIONS
PERSON	DIVING	SENSE	SHORT	ELECTRIC	**FIELD**	PLAYING	FIND
INFECTIOUS	DOLPHINS	CONSCIOUSNESS	FICTION	DIRECTION	PHYSICS	HIT	POSITION
COMMON	SWAM	STRANGE	ACTION	FORCE	LABORATORY	TENNIS	**FIELD**
CAUSING	LONG	FEELING	TRUE	MAGNETS	STUDIES	TEAMS	OCCUPATIONS
SMALLPOX	SEAL	WHOLE	EVENTS	BE	WORLD	GAMES	REQUIRE
BODY	DIVE	BEING	TELLS	MAGNETISM	SCIENTIST	SPORTS	OPPORTUNITY
INFECTIONS	DOLPHIN	MIGHT	TALE	POLE	STUDYING	BAT	EARN
CERTAIN	UNDERWATER	HOPE	NOVEL	INDUCED	SCIENCES	TERRY	ABLE

[Tennenbaum et al]

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

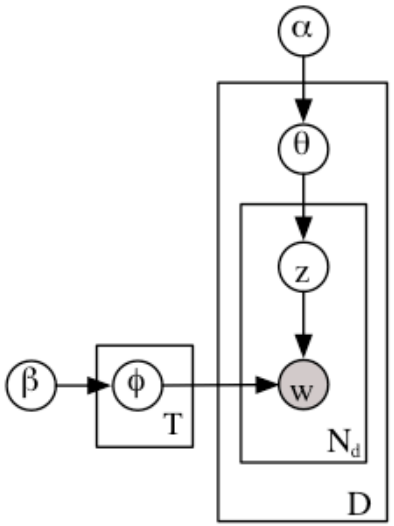
From LDA to Author-Recipient-Topic

(ART)

Latent Dirichlet Allocation

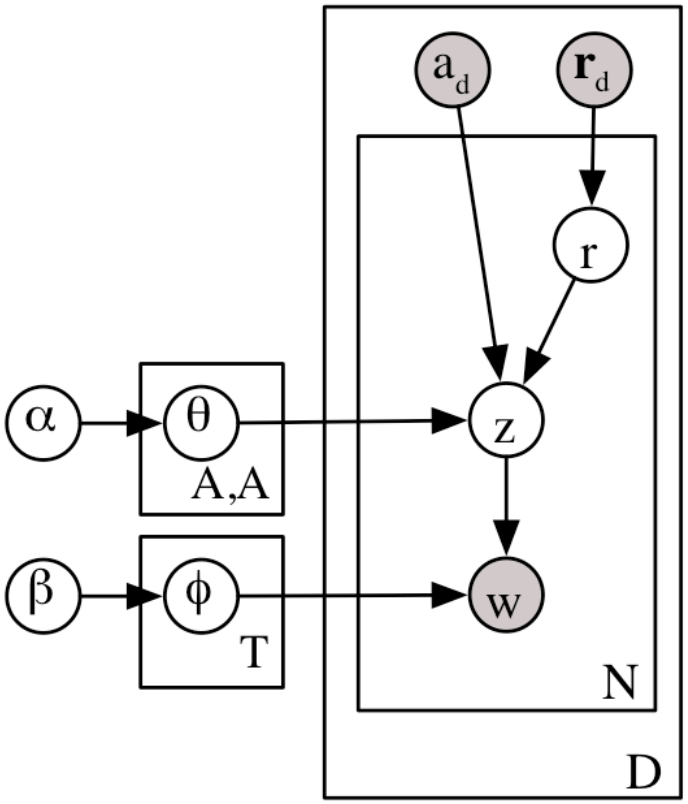
(LDA)

[Blei, Ng, Jordan, 2003]



Inference and Estimation

$$p(\theta, \phi, \mathbf{x}_d, \mathbf{z}_d, \mathbf{w}_d | \alpha, \beta, a_d, \mathbf{r}_d) = p(\theta|\alpha)p(\phi|\beta) \prod_{n=1}^{N_d} p(x_{dn}|\mathbf{r}_d)p(z_{dn}|\theta_{a_d,x_{dn}})p(w_{dn}|\phi_{z_{dn}})$$



Gibbs Sampling:

- Easy to implement
- Reasonably fast

$$P(z_i \,|\, \mathbf{z}_{-i}, \mathbf{x}, \mathbf{w}) \quad \propto \quad \frac{n_{z_i}^{w_v} + \beta_v}{\sum_v n_{z_i}^{w_v} + \beta_v} \frac{n_{x_i}^{z_i} + \alpha_{z_i}}{\sum_{z'} n_{x_i}^{z'} + \alpha_{z'}}$$

$$P(\mathrm{r}_i \,|\, \mathbf{z}, \mathrm{r}_{-i}, \mathbf{w}) \quad \propto \quad \frac{n_{x_i}^{z_i} + \alpha_{z_i}}{\sum_{z'} n_{x_i}^{z'} + \alpha_{z'}}$$

Enron Email Corpus

- 250k email messages
- 23k people

```
Date: Wed, 11 Apr 2001 06:56:00 -0700 (PDT)
From: debra.perlingiere@enron.com
To: steve.hooser@enron.com
Subject: Enron/TransAltaContract dated Jan 1, 2001

Please see below. Katalin Kiss of TransAlta has requested an
electronic copy of our final draft?  Are you OK with this?  If
so, the only version I have is the original draft without
revisions.

DP

Debra Perlingiere
Enron North America Corp.
Legal Department
1400 Smith Street, EB 3885
Houston, Texas 77002
dperlin@enron.com
```

Topics, and prominent senders / receivers discovered by ART

Topic names, by hand

Topic 5 "Legal Contracts"		Topic 17 "Document Review"		Topic 27 "Time Scheduling"		Topic 45 "Sports Pool"	
section	0.0299	attached	0.0742	day	0.0419	game	0.0170
party	0.0265	agreement	0.0493	friday	0.0418	draft	0.0156
language	0.0226	review	0.0340	morning	0.0369	week	0.0135
contract	0.0203	questions	0.0257	monday	0.0282	team	0.0135
date	0.0155	draft	0.0245	office	0.0282	eric	0.0130
enron	0.0151	letter	0.0239	wednesday	0.0267	make	0.0125
parties	0.0149	comments	0.0207	tuesday	0.0261	free	0.0107
notice	0.0126	copy	0.0165	time	0.0218	year	0.0106
days	0.0112	revised	0.0161	good	0.0214	pick	0.0097
include	0.0111	document	0.0156	thursday	0.0191	phillip	0.0095
M.Hain J.Steffes	0.0549	G.Nemec B.Tycholiz	0.0737	J.Dasovich R.Shapiro	0.0340	E.Bass M.Lenhart	0.3050
J.Dasovich R.Shapiro	0.0377	G.Nemec M.Whitt	0.0551	J.Dasovich J.Steffes	0.0289	E.Bass P.Love	0.0780
D.Hyvl K.Ward	0.0362	B.Tycholiz G.Nemec	0.0325	C.Clair M.Taylor	0.0175	M.Motley M.Grigsby	0.0522

Topics, and prominent senders / receivers discovered by ART

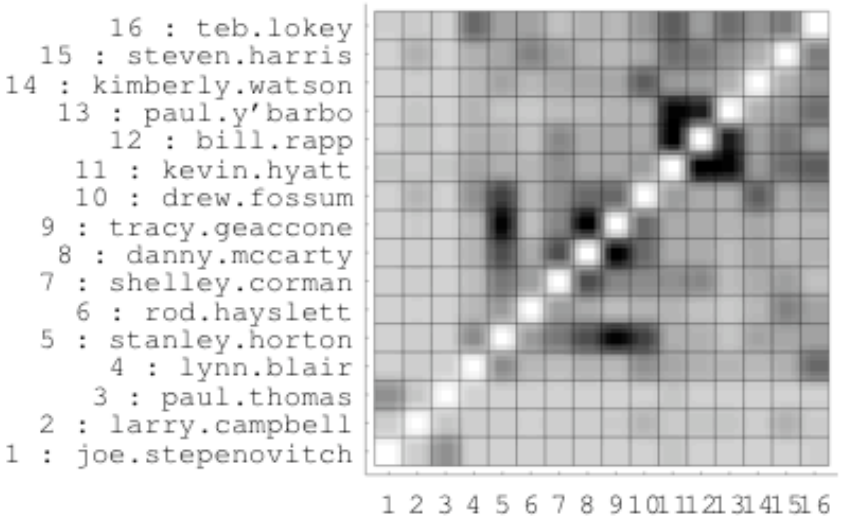
Topic 34 "Operations"		Topic 37 "Power Market"		Topic 41 "Government Relations"		Topic 42 "Wireless"	
operations	0.0321	market	0.0567	state	0.0404	blackberry	0.0726
team	0.0234	power	0.0563	california	0.0367	net	0.0557
office	0.0173	price	0.0280	power	0.0337	www	0.0409
list	0.0144	system	0.0206	energy	0.0239	website	0.0375
bob	0.0129	prices	0.0182	electricity	0.0203	report	0.0373
open	0.0126	high	0.0124	davis	0.0183	wireless	0.0364
meeting	0.0107	based	0.0120	utilities	0.0158	handheld	0.0362
gas	0.0107	buy	0.0117	commission	0.0136	stan	0.0282
business	0.0106	customers	0.0110	governor	0.0132	fyi	0.0271
houston	0.0099	costs	0.0106	prices	0.0089	named	0.0260
S.Beck L.Kitchen	0.2158	J.Dasovich J.Steffes	0.1231	J.Dasovich R.Shapiro	0.3338	R.Haylett T.Geaccone	0.1432
S.Beck J.Lavorato	0.0826	J.Dasovich R.Shapiro	0.1133	J.Dasovich J.Steffes	0.2440	T.Geaccone R.Haylett	0.0737
S.Beck S.White	0.0530	M.Taylor E.Sager	0.0218	J.Dasovich R.Sanders	0.1394	R.Haylett D.Fossum	0.0420

Beck = "Chief Operations Officer"

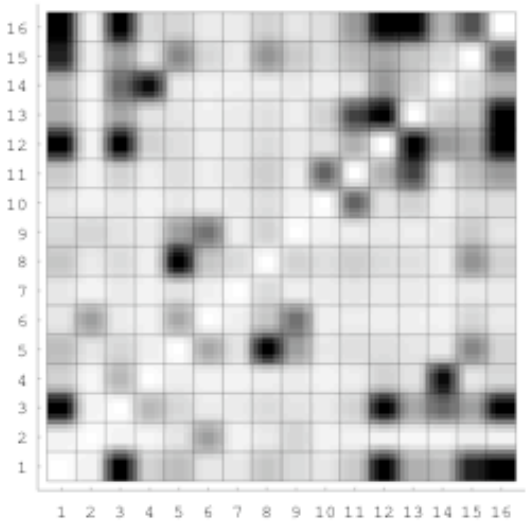
Dasovich = "Government Relations Executive"
Shapiro = "Vice President of Regulatory Affairs"
Steffes = "Vice President of Government Affairs"

Comparing Role Discovery

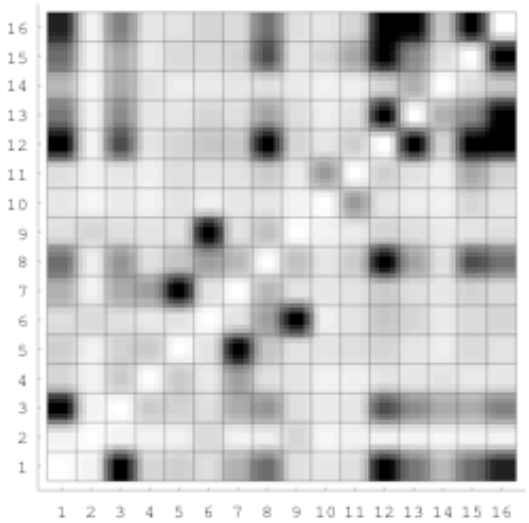
Traditional SNA



ART



Author-Topic



connection strength (A,B) =

distribution over recipients

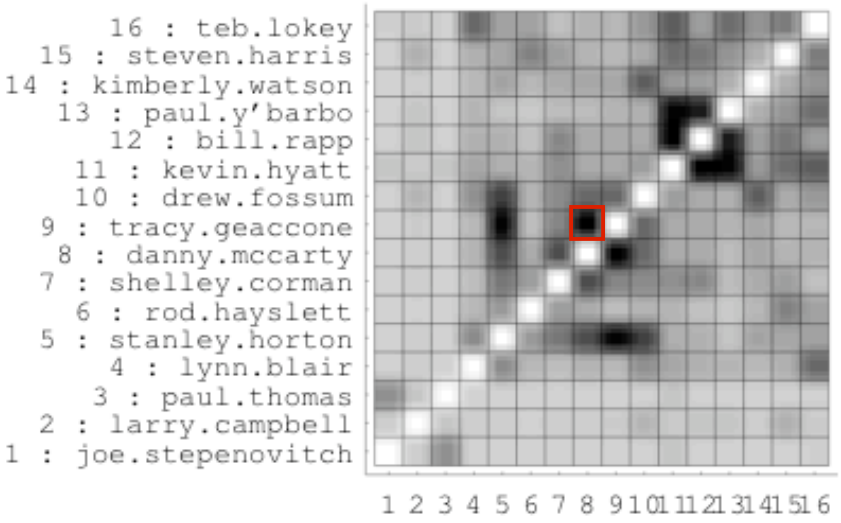
distribution over authored topics

distribution over authored topics

Comparing Role Discovery

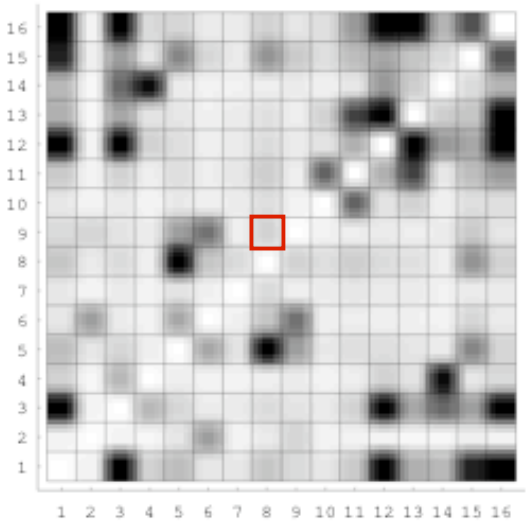
Tracy Geaconne ⇔ Dan McCarty

Traditional SNA



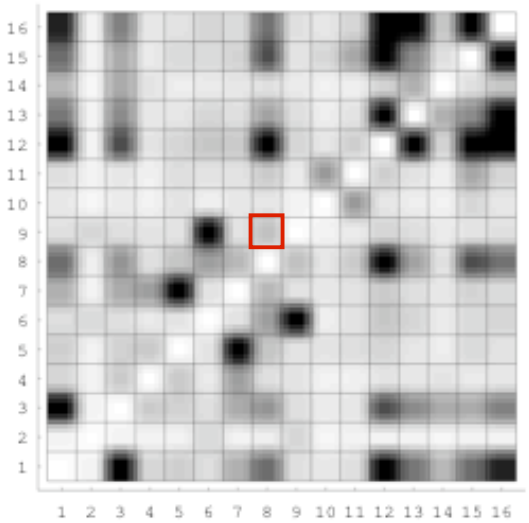
Similar roles

ART



Different roles

Author-Topic



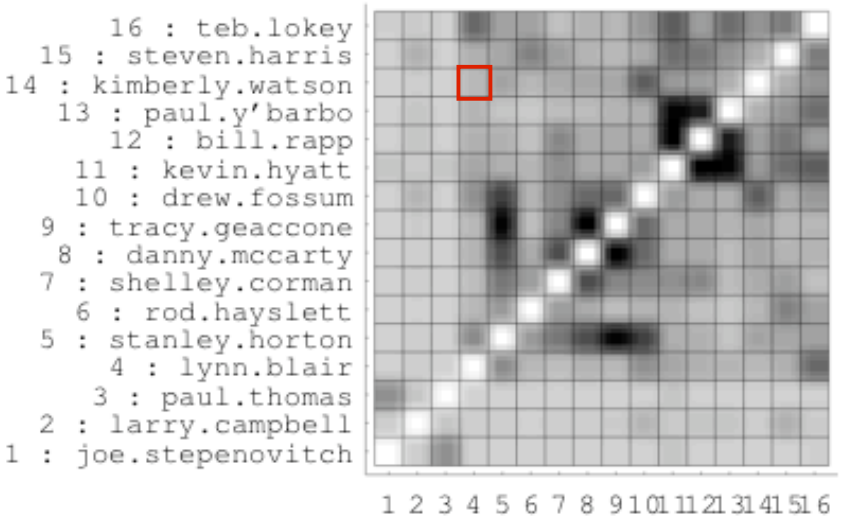
Different roles

Geaconne = "Secretary"
McCarty = "Vice President"

Comparing Role Discovery

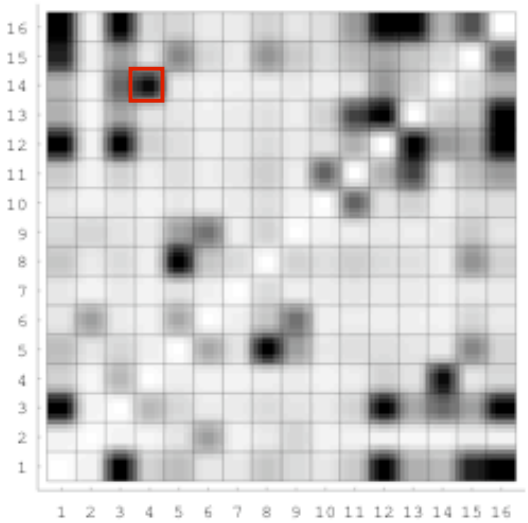
Lynn Blair ⇔ Kimberly Watson

Traditional SNA



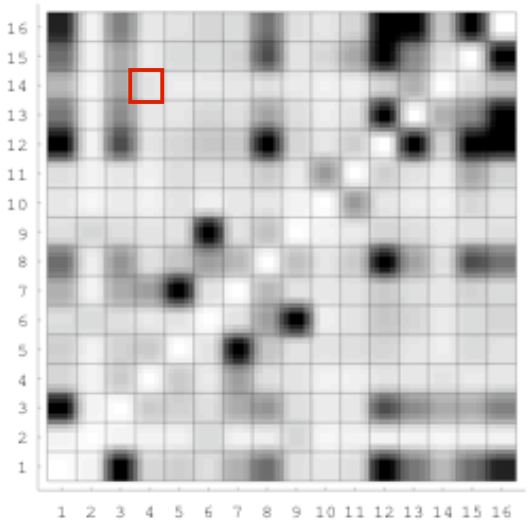
Different roles

ART



Very similar

Author-Topic



Very different

Blair = "Gas pipeline logistics"
Watson = "Pipeline facilities planning"

McCallum Email Corpus 2004

- January - October 2004
- 23k email messages
- 825 people

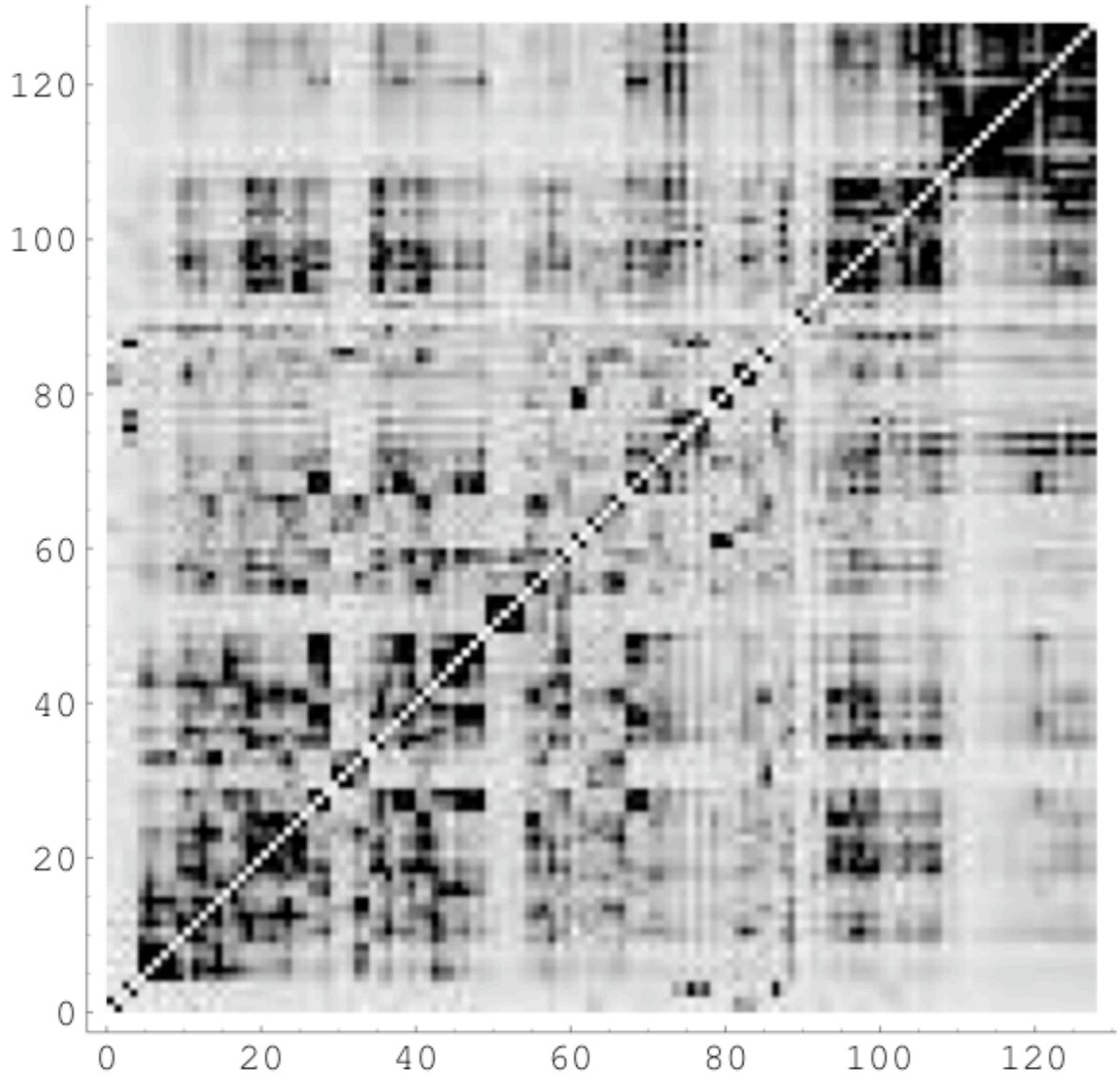
```
From: kate@cs.umass.edu
Subject: NIPS and ....
Date: June 14, 2004 2:27:41 PM EDT
To: mccallum@cs.umass.edu

There is pertinent stuff on the first yellow folder that is
completed either travel or other things, so please sign that
first folder anyway. Then, here is the reminder of the things
I'm still waiting for:

NIPS registration receipt.
CALO registration receipt.

Thanks,
Kate
```

McCallum Email Blockstructure



Four most prominent topics in discussions with ____?

Topic 5 "Grant Proposals"		Topic 31 "Meeting Setup"		Topic 38 "ML Models"		Topic 41 "Friendly Discourse"	
proposal	0.0397	today	0.0512	model	0.0479	great	0.0516
data	0.0310	tomorrow	0.0454	models	0.0444	good	0.0393
budget	0.0289	time	0.0413	inference	0.0191	don	0.0223
work	0.0245	ll	0.0391	conditional	0.0181	sounds	0.0219
year	0.0238	meeting	0.0339	methods	0.0144	work	0.0196
glenn	0.0225	week	0.0255	number	0.0136	wishes	0.0182
nsf	0.0209	talk	0.0246	sequence	0.0126	talk	0.0175
project	0.0188	meet	0.0233	learning	0.0126	interesting	0.0168
sets	0.0157	morning	0.0228	graphical	0.0121	time	0.0162
support	0.0156	monday	0.0208	random	0.0121	hear	0.0132

Topic 5 "Grant Proposals"		Topic 31 "Meeting Setup"		Topic 38 "ML Models"		Topic 41 "Friendly Discourse"	
proposal	0.0397	today	0.0512	model	0.0479	great	0.0516
data	0.0310	tomorrow	0.0454	models	0.0444	good	0.0393
budget	0.0289	time	0.0413	inference	0.0191	don	0.0223
work	0.0245	ll	0.0391	conditional	0.0181	sounds	0.0219
year	0.0238	meeting	0.0339	methods	0.0144	work	0.0196
glenn	0.0225	week	0.0255	number	0.0136	wishes	0.0182
nsf	0.0209	talk	0.0246	sequence	0.0126	talk	0.0175
project	0.0188	meet	0.0233	learning	0.0126	interesting	0.0168
sets	0.0157	morning	0.0228	graphical	0.0121	time	0.0162
support	0.0156	monday	0.0208	random	0.0121	hear	0.0132
smyth mccallum	0.1290	ronb mccallum	0.0339	casutton mccallum	0.0498	mccallum culotta	0.0558
mccallum stowell	0.0746	wellner mccallum	0.0314	icml04-webadmin icml04-chairs	0.0366	mccallum casutton	0.0530
mccallum lafferty	0.0739	casutton mccallum	0.0217	mccallum casutton	0.0343	mccallum ronb	0.0274
mccallum smyth	0.0532	mccallum casutton	0.0200	nips04workflow mccallum	0.0322	mccallum saunders	0.0255
pereira lafferty	0.0339	mccallum wellner	0.0200	weinman mccallum	0.0250	mccallum pereira	0.0181

Two most prominent topics in discussions with ____?

Topic 1

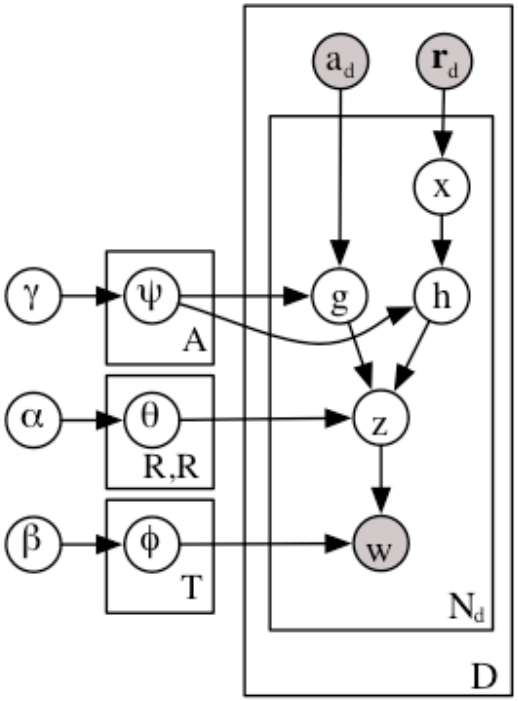
Words	Prob
love	0.030514
house	0.015402
	0.013659
time	0.012351
great	0.011334
hope	0.011043
dinner	0.00959
saturday	0.009154
left	0.009154
ll	0.009009
	0.008282
visit	0.008137
evening	0.008137
stay	0.007847
bring	0.007701
weekend	0.007411
road	0.00712
sunday	0.006829
kids	0.006539
flight	0.006539

Topic 2

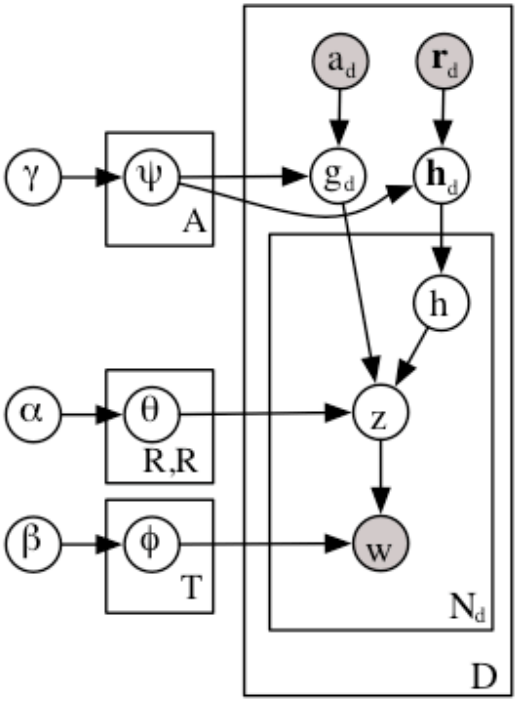
Words	Prob
today	0.051152
tomorrow	0.045393
time	0.041289
ll	0.039145
meeting	0.033877
week	0.025484
talk	0.024626
meet	0.023279
morning	0.022789
monday	0.020767
back	0.019358
call	0.016418
free	0.015621
home	0.013967
won	0.013783
day	0.01311
hope	0.012987
leave	0.012987
office	0.012742
tuesday	0.012558

Role-Author-Recipient-Topic Models

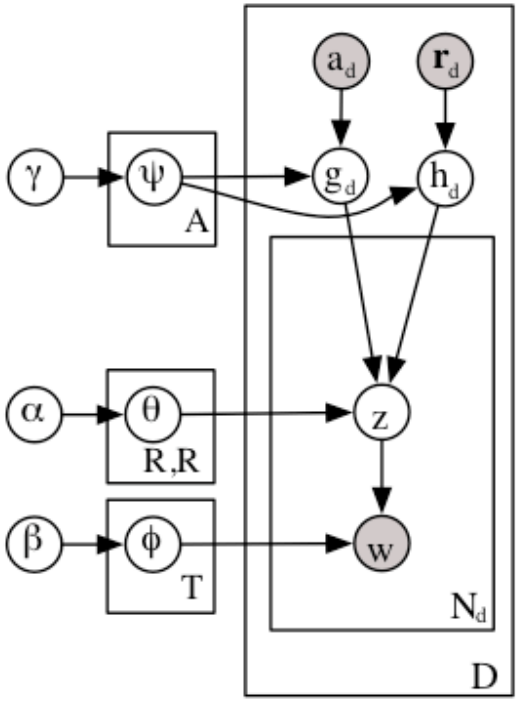
Role-Author-Recipient-Topic Model 1
(RART1)



Role-Author-Recipient-Topic Model 2
(RART2)



Role-Author-Recipient-Topic Model 3
(RART3)



Results with RART: People in “Role #3” in Academic Email

- **`olc`** lead Linux sysadmin
- **`gauthier`** sysadmin for CIIR group
- **`irsystem`** mailing list CIIR sysadmins
- **`system`** mailing list for dept. sysadmins
- **`allan`** Prof., chair of “computing committee”
- **`valerie`** second Linux sysadmin
- **`tech`** mailing list for dept. hardware
- **`steve`** head of dept. I.T. support

Roles for `allan` (James Allan)

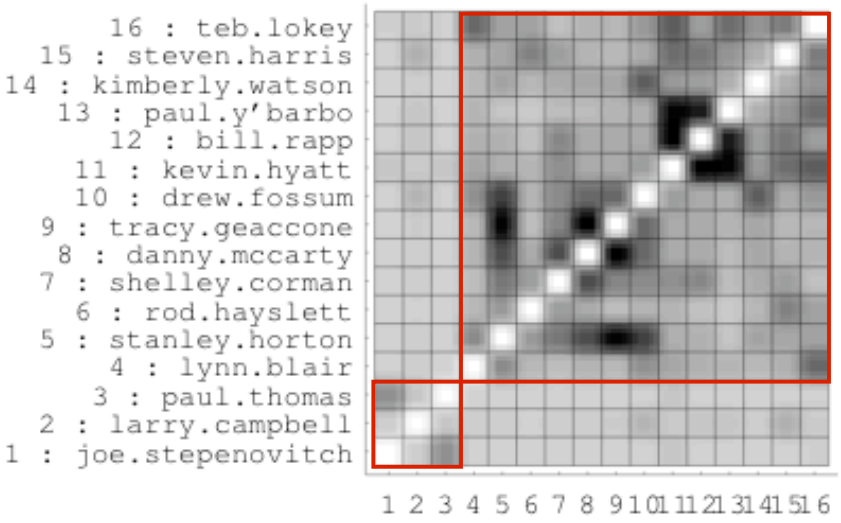
- Role #3 I.T. support
- Role #2 Natural Language researcher

Roles for `pereira` (Fernando Pereira)

- Role #2 Natural Language researcher
- Role #4 SRI CALO project participant
- Role #6 Grant proposal writer
- Role #10 Grant proposal coordinator
- Role #8 Guests at McCallum's house

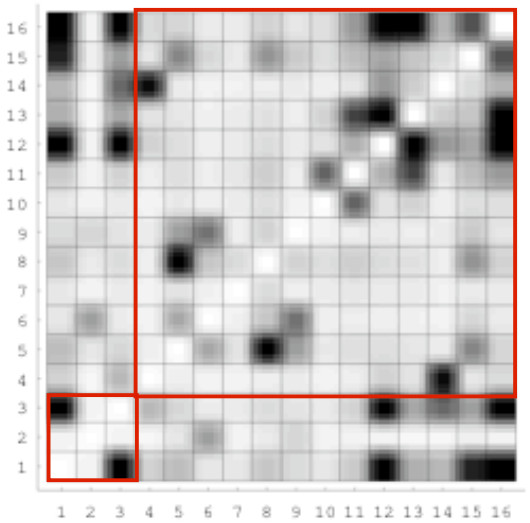
ART: Roles but not Groups

Traditional SNA



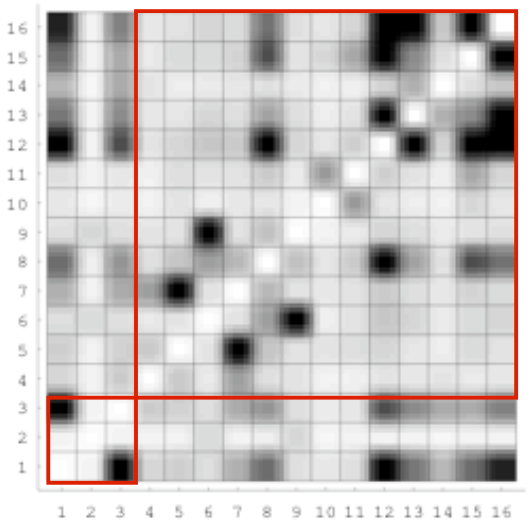
Block structured

ART



Not

Author-Topic



Not

Enron TransWestern Division

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

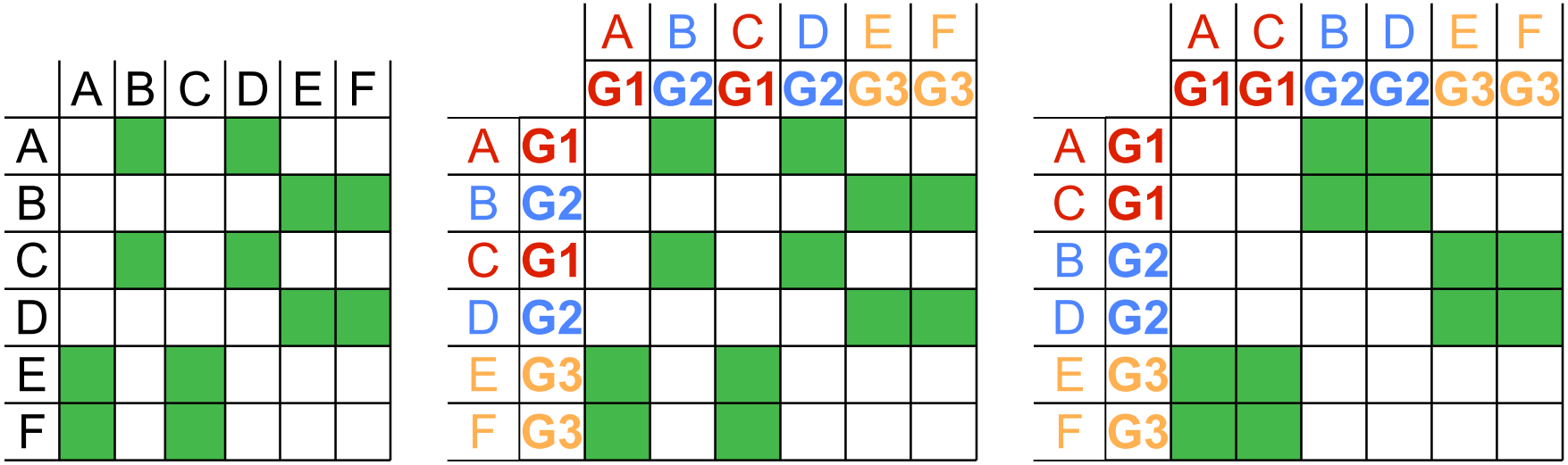
Groups and Topics

- Input:
  - Observed relations between people
  - Attributes on those relations (text, or categorical)

- Output:
  - Attributes clustered into “topics”
  - Groups of people---varying depending on topic

Adjacency Matrix Representing Relations

Student Roster	Academic Admiration
Adams	Acad(A, B) Acad(C, B)
Bennett	Acad(A, D) Acad(C, D)
Carter	Acad(B, E) Acad(D, E)
Davis	Acad(B, F) Acad(D, F)
Edwards	Acad(E, A) Acad(F, A)
Frederking	Acad(E, C) Acad(F, C)

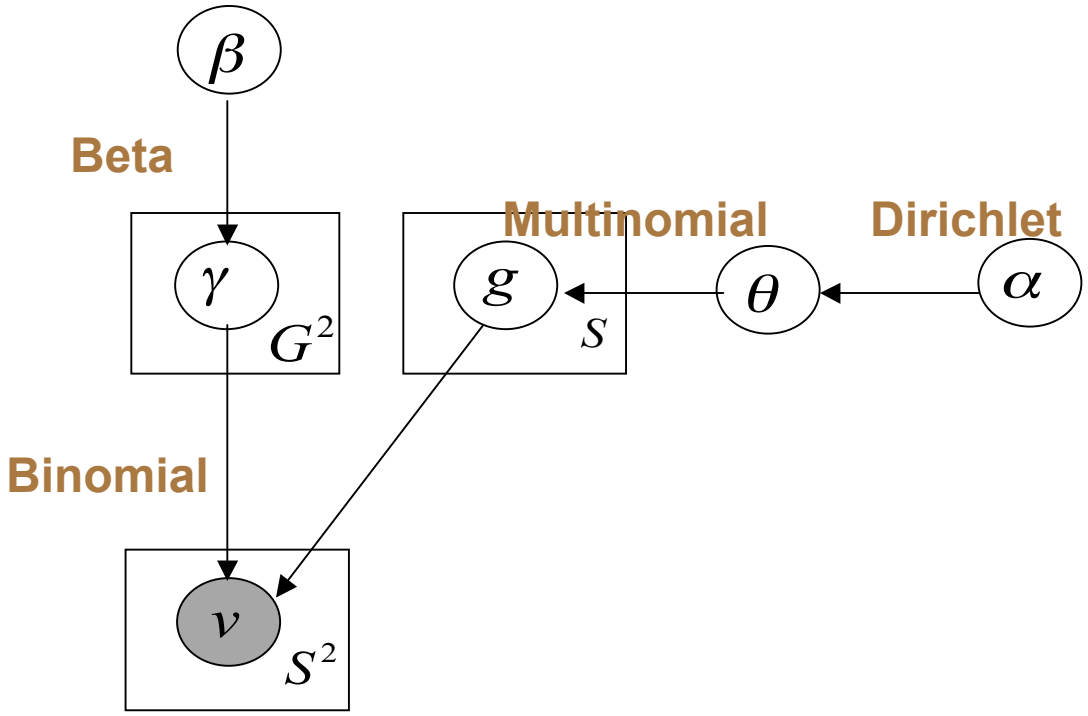


Group Model: Partitioning Entities into Groups

Stochastic Blockstructures for Relations
[Nowicki, Snijders 2001]

S: number of entities

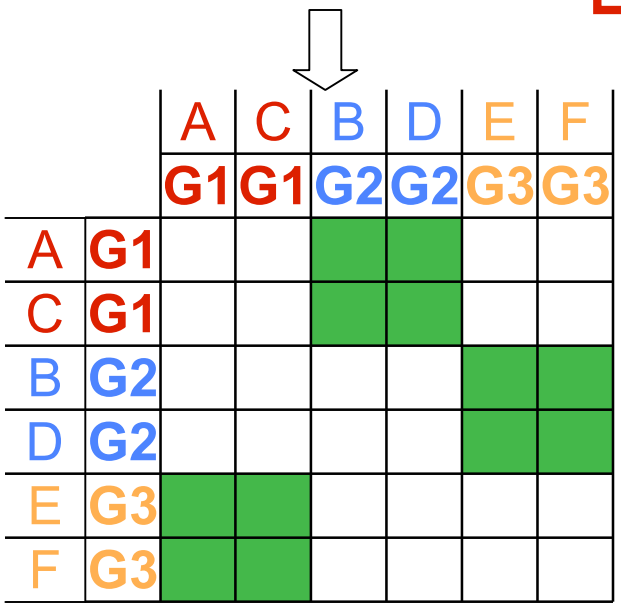
G: number of groups

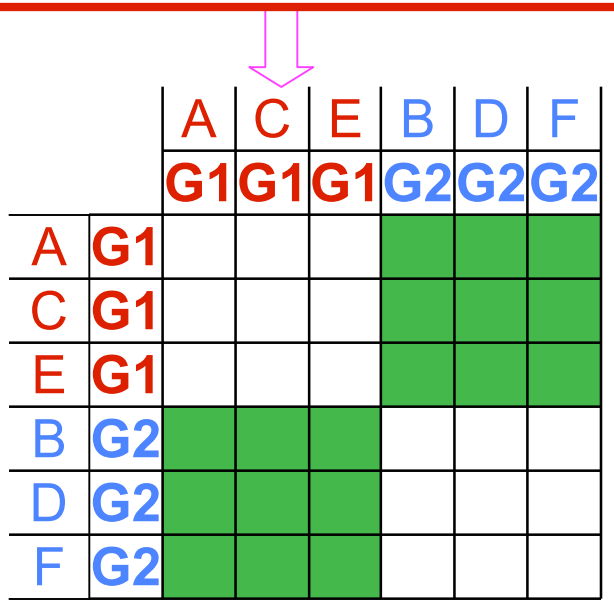


Enhanced with arbitrary number of groups in [Kemp, Griffiths, Tenenbaum 2004]

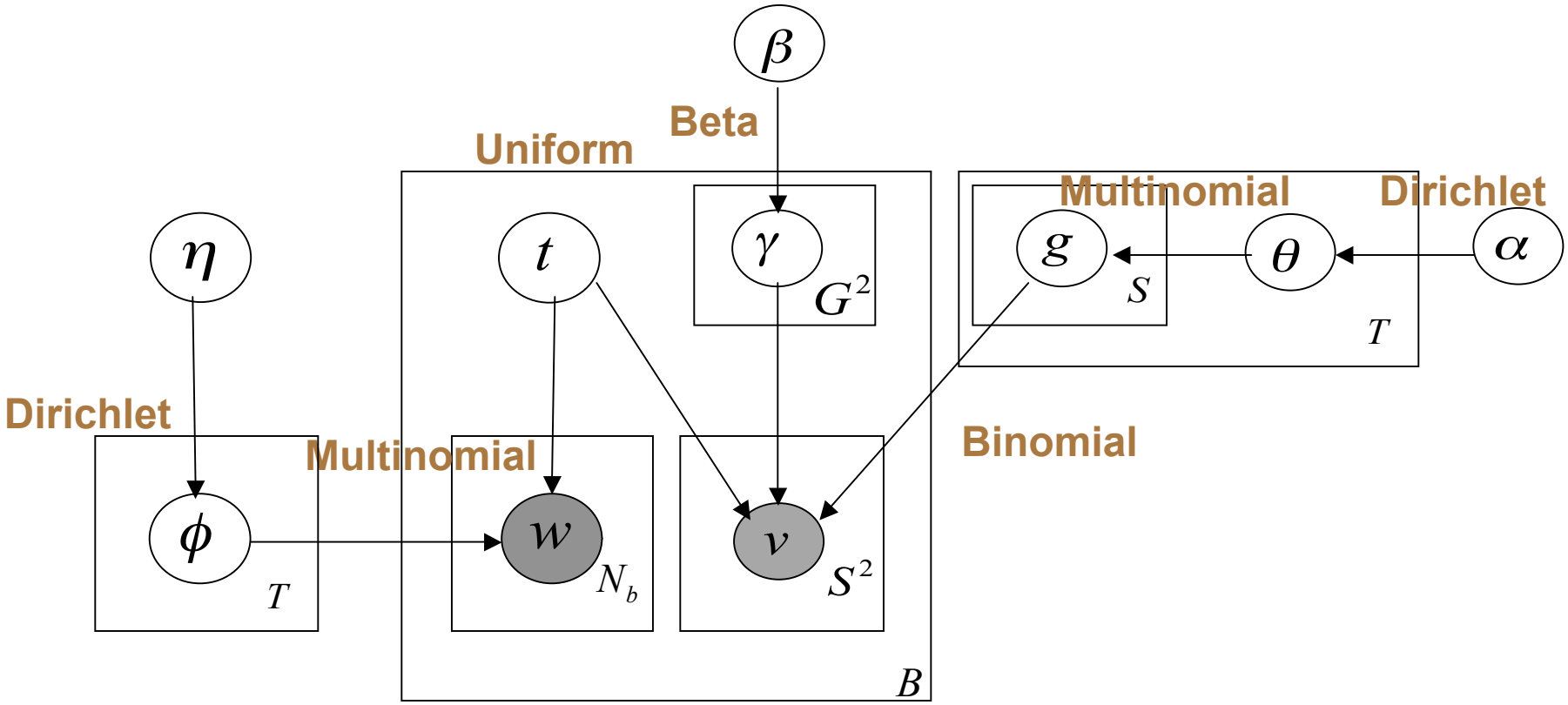
Two Relations with Different Attributes

Student Roster	Academic Admiration	Social Admiration
Adams	Acad(A, B) Acad(C, B)	Soci(A, B) Soci(A, D) Soci(A, F)
Bennett	Acad(A, D) Acad(C, D)	Soci(B, A) Soci(B, C) Soci(B, E)
Carter	Acad(B, E) Acad(D, E)	Soci(C, B) Soci(C, D) Soci(C, F)
Davis	Acad(B, F) Acad(D, F)	Soci(D, A) Soci(D, C) Soci(D, E)
Edwards	Acad(E, A) Acad(F, A)	Soci(E, B) Soci(E, D) Soci(E, F)
Frederking	Acad(E, C) Acad(F, C)	Soci(F, A) Soci(F, C) Soci(F, E)





The Group-Topic Model: Discovering Groups and Topics Simultaneously



Dataset #1:
U.S. Senate

- **16 years of voting records in the US Senate (1989 – 2005)**
- **a Senator may respond *Yea* or *Nay* to a resolution**
- **3423 resolutions with text attributes (index terms)**
- **191 Senators in total across 16 years**

S.543
Title: An Act to reform Federal deposit insurance, protect the deposit insurance funds, recapitalize the Bank Insurance Fund, improve supervision and regulation of insured depository institutions, and for other purposes.
Sponsor: Sen Riegle, Donald W., Jr. [MI] (introduced 3/5/1991) Cosponsors (2)
Latest Major Action: 12/19/1991 Became Public Law No: 102-242.
Index terms: Banks and banking Accounting Administrative fees Cost control Credit Deposit insurance Depressed areas and other 110 terms

Adams (D-WA), **Nay** Akaka (D-HI), **Yea** Bentsen (D-TX), **Yea** Biden (D-DE), **Yea** Bond (R-MO), **Yea** Bradley (D-NJ), **Nay** Conrad (D-ND), **Nay ……**

Topics Discovered (U.S. Senate)

Mixture of Unigrams

Education	Energy	Military Misc.	Economic
education	energy	government	federal
school	power	military	labor
aid	water	foreign	insurance
children	nuclear	tax	aid
drug	gas	congress	tax
students	petrol	aid	business
elementary	research	law	employee
prevention	pollution	policy	care

Group-Topic Model

Education + Domestic	Foreign	Economic	Social Security + Medicare
education	foreign	labor	social
school	trade	insurance	security
federal	chemicals	tax	insurance
aid	tariff	congress	medical
government	congress	income	care
tax	drugs	minimum	medicare
energy	communicable	wage	disability
research	diseases	business	assistance

Groups Discovered (US Senate)

Groups from topic ***Education + Domestic***

Group 1	Group 3	Group 4
73 Republicans	Cohen(R-ME)	Armstrong(R-CO)
Krueger(D-TX)	Danforth(R-MO)	Garn(R-UT)
Group 2	Durenberger(R-MN)	Humphrey(R-NH)
90 Democrats	Hatfield(R-OR)	McCain(R-AZ)
Chafee,L.(R-RI)	Heinz(R-PA)	McClure(R-ID)
Jeffords(I-VT)	Jeffords(R-VT)	Roth(R-DE)
	Kassebaum(R-KS)	Symms(R-ID)
	Packwood(R-OR)	Wallop(R-WY)
	Specter(R-PA)	Brown(R-CO)
	Snowe(R-ME)	DeWine(R-OH)
	Collins(R-ME)	Thompson(R-TN)
		Fitzgerald(R-IL)
		Voinovich(R-OH)
		Miller(D-GA)
		Coleman(R-MN)

Senators Who Change Coalition the most Dependent on Topic

Senator	Group Switch Index
Shelby(D-AL)	0.6182
Heflin(D-AL)	0.6049
Voinovich(R-OH)	0.6012
Johnston(D-LA)	0.5878
Armstrong(R-CO)	0.5747

e.g. Senator Shelby (D-AL) votes
with the Republicans on Economic
with the Democrats on Education + Domestic
with a small group of maverick Republicans on Social Security + Medicaid

Dataset #2:

The UN General Assembly

- Voting records of the UN General Assembly (1990 - 2003)
- A country may choose to vote *Yes*, *No* or *Abstain*
- 931 resolutions with text attributes (titles)
- 192 countries in total
- Also experiments later with resolutions from 1960-2003

Vote on Permanent Sovereignty of Palestinian People, 87th plenary meeting

The draft resolution on permanent sovereignty of the Palestinian people in the occupied Palestinian territory, including Jerusalem, and of the Arab population in the occupied Syrian Golan over their natural resources (document A/54/591) was adopted by a recorded vote of 145 in favour to 3 against with 6 abstentions:

In favour: Afghanistan, Argentina, Belgium, Brazil, Canada, China, France, Germany, India, Japan, Mexico, Netherlands, New Zealand, Pakistan, Panama, Russian Federation, South Africa, Spain, Turkey, and other 126 countries.
Against: Israel, Marshall Islands, United States.
Abstain: Australia, Cameroon, Georgia, Kazakhstan, Uzbekistan, Zambia.

Topics Discovered (UN)

Mixture of Unigrams

Everything Nuclear	Human Rights	Security in Middle East
nuclear	rights	occupied
weapons	human	israel
use	palestine	syria
implementation	situation	security
countries	israel	calls

Group-Topic Model

Nuclear Non-proliferation	Nuclear Arms Race	Human Rights
nuclear	nuclear	rights
states	arms	human
united	prevention	palestine
weapons	race	occupied
nations	space	israel

Groups Discovered (UN)

The countries list for each group are ordered by their 2005 GDP (PPP) and only 5 countries are shown in groups that have more than 5 members.

GROUP ↓	Nuclear Arsenal	Human Rights	Nuclear Arms Race
	nuclear	rights	nuclear
	states	human	arms
	united	palestine	prevention
	weapons	occupied	race
	nations	israel	space
1	Brazil	Brazil	UK
	Columbia	Mexico	France
	Chile	Columbia	Spain
	Peru	Chile	Monaco
	Venezuela	Peru	East-Timor
2	USA	Nicaragua	India
	Japan	Papua	Russia
	Germany	Rwanda	Micronesia
	UK...	Swaziland	
	Russia	Fiji	
3	China	USA	Japan
	India	Japan	Germany
	Mexico	Germany	Italy...
	Iran	UK...	Poland
	Pakistan	Russia	Hungary
4	Kazakhstan	China	China
	Belarus	India	Brazil
	Yugoslavia	Indonesia	Mexico
	Azerbaijan	Thailand	Indonesia
	Cyprus	Philippines	Iran
5	Thailand	Belarus	USA
	Philippines	Turkmenistan	Israel
	Malaysia	Azerbaijan	Palau
	Nigeria	Uruguay	
	Tunisia	Kyrgyzstan	

Groups and Topics, Trends over Time (UN)

Time Period	Topic 1	Topic 2	Topic 3	Group distributions for Topic 3				
				Group 1	Group2	Group3	Group4	Group5
60-75	Nuclear	Procedure	Africa Indep.	India	USA	Argentina	USSR	Turkey
	operative	committee	calling	Indonesia	Japan	Colombia	Poland	
	general	amendment	right	Iran	UK	Chile	Hungary	
	nuclear	assembly	africa	Thailand	France	Venezuela	Bulgaria	
	power	deciding	self	Philippines	Italy	Dominican	Belarus	
65-80	Independence	Finance	Weapons	Cuba	India	Algeria	USSR	USA
	territories	budget	nuclear	Albania	Indonesia	Iraq	Poland	Japan
	independence	appropriation	UN		Pakistan	Syria	Hungary	UK
	self	contribution	international		Saudi	Libya	Bulgaria	France
	colonial	income	weapons		Egypt	Afganistan	Belarus	Italy
70-85	N. Weapons	Israel	Rights	Mexico	China	USA	Brazil	India
	nuclear	israel	africa	Indonesia		Japan	Turkey	USSR
	international	measures	territories	Iran		UK	Argentina	Poland
	UN	hebron	south	Thailand		France	Colombia	Vietnam
	human	expelling	right	Philippines		Italy	Chile	Hungary
75-90	Rights	Israel/Pal.	Disarmament	Mexico	USA	Algeria	China	India
	south	israel	UN	Indonesia	Japan	Vietnam	Brazil	
	africa	arab	international	Iran	UK	Iraq	Argentina	
	israel	occupied	nuclear	Thailand	France	Syria	Colombia	
	rights	palestine	disarmament	Philippines	USSR	Libya	Chile	
80-95	Disarmament	Conflict	Pal. Rights	USA	China	Japan	Guatemala	Malawi
	nuclear	need	rights	Israel	India	UK	St Vincent	
	US	israel	palestine		Russia	France	Dominican	
	disarmament	palestine	israel		Spain	Italy		
	international	secretary	occupied		Hungary	Canada		
85-00	Weapons	Rights	Israel/Pal.	Poland	China	USA	Russia	Cameroon
	nuclear	rights	israeli	Czech R.	India	Japan	Argentina	Congo
	weapons	human	palestine	Hungary	Brazil	UK	Ukraine	Ivory C.
	use	fundamental	occupied	Bulgaria	Mexico	France	Belarus	Liberia
	international	freedoms	disarmament	Albania	Indonesia	Italy	Malta	

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Path Analysis *(Topical Sequence Model, TSM)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

Groups and Topics, Trends over Time (UN)

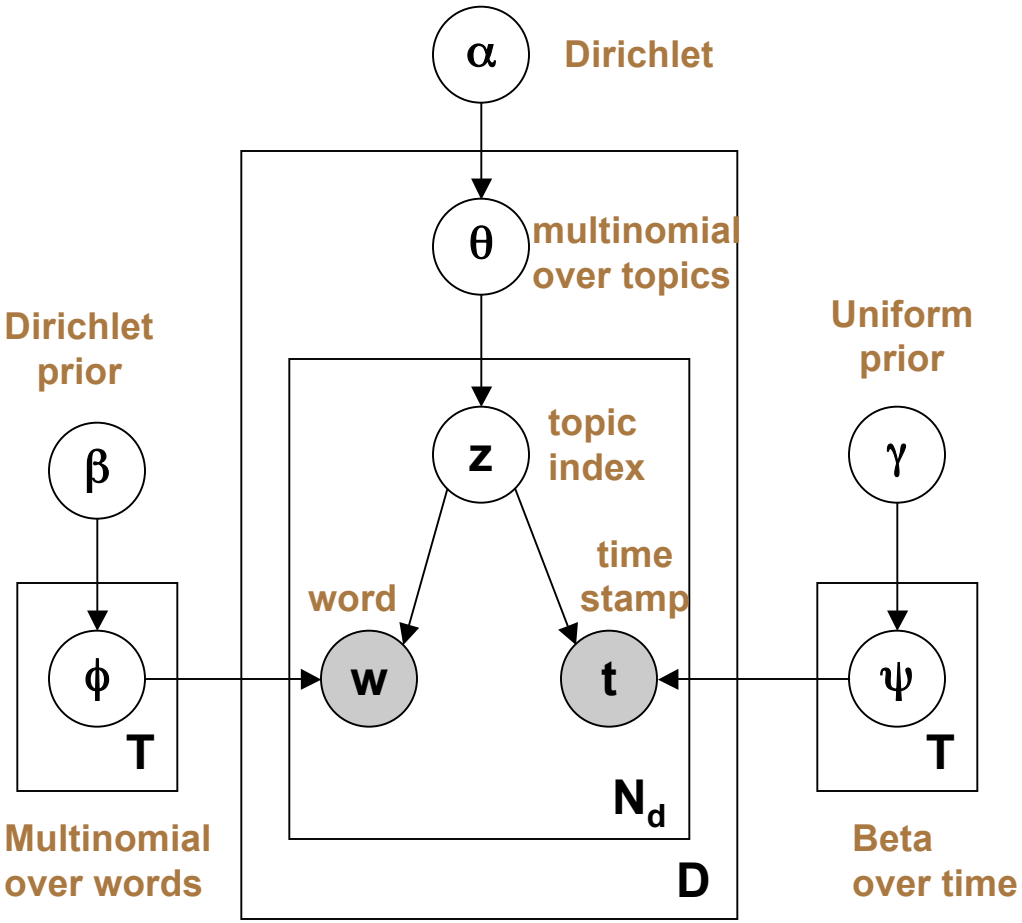
Time Period	Topic 1	Topic 2	Topic 3	Group distributions for Topic 3				
				Group 1	Group2	Group3	Group4	Group5
60-75	Nuclear	Procedure	Africa Indep.	India	USA	Argentina	USSR	Turkey
	operative	committee	calling	Indonesia	Japan	Colombia	Poland	
	general	amendment	right	Iran	UK	Chile	Hungary	
	nuclear	assembly	africa	Thailand	France	Venezuela	Bulgaria	
	power	deciding	self	Philippines	Italy	Dominican	Belarus	
65-80	Independence	Finance	Weapons	Cuba	India	Algeria	USSR	USA
	territories	budget	nuclear	Albania	Indonesia	Iraq	Poland	Japan
	independence	appropriation	UN		Pakistan	Syria	Hungary	UK
	self	contribution	international		Saudi	Libya	Bulgaria	France
	colonial	income	weapons		Egypt	Afganistan	Belarus	Italy
70-85	N. Weapons	Israel	Rights	Mexico	China	USA	Brazil	India
	nuclear	israel	africa	Indonesia		Japan	Turkey	USSR
	international	measures	territories	Iran		UK	Argentina	Poland
	UN	hebron	south	Thailand		France	Colombia	Vietnam
	human	expelling	right	Philippines		Italy	Chile	Hungary
75-90	Rights	Israel/Pal.	Disarmament	Mexico	USA	Algeria	China	India
	south	israel	UN	Indonesia	Japan	Vietnam	Brazil	
	africa	arab	international	Iran	UK	Iraq	Argentina	
	israel	occupied	nuclear	Thailand	France	Syria	Colombia	
	rights	palestine	disarmament	Philippines	USSR	Libya	Chile	
80-95	Disarmament	Conflict	Pal. Rights	USA	China	Japan	Guatemala	Malawi
	nuclear	need	rights	Israel	India	UK	St Vincent	
	US	israel	palestine		Russia	France	Dominican	
	disarmament	palestine	israel		Spain	Italy		
	international	secretary	occupied		Hungary	Canada		
85-00	Weapons	Rights	Israel/Pal.	Poland	China	USA	Russia	Cameroon
	nuclear	rights	israeli	Czech R.	India	Japan	Argentina	Congo
	weapons	human	palestine	Hungary	Brazil	UK	Ukraine	Ivory C.
	use	fundamental	occupied	Bulgaria	Mexico	France	Belarus	Liberia
	international	freedoms	disarmament	Albania	Indonesia	Italy	Malta	

Want to Model Trends over Time

- Pattern appears only briefly
  - Capture its statistics in focused way
  - Don’t confuse it with patterns elsewhere in time
- Is prevalence of topic growing or waning?
- How do roles, groups, influence shift over time?

Topics over Time (TOT)

[Wang, McCallum, KDD 2006]



State of the Union Address

208 Addresses delivered between January 8, 1790 and January 29, 2002.

To increase the number of documents, we split the addresses into paragraphs and treated them as ‘documents’. One-line paragraphs were excluded. Stopping was applied.

- **17156 ‘documents’**
- **21534 words**
- **669,425 tokens**

> Our scheme of taxation, by means of which this needless surplus is taken from the people and put into the public Treasury, consists of a tariff or duty levied upon importations from abroad and internal-revenue taxes levied upon the consumption of tobacco and spirituous and malt liquors. It must be conceded that none of the things subjected to internal-revenue taxation are, strictly speaking, necessaries. There appears to be no just complaint of this taxation by the consumers of these articles, and there seems to be nothing so well able to bear the burden without hardship to any portion of the people.

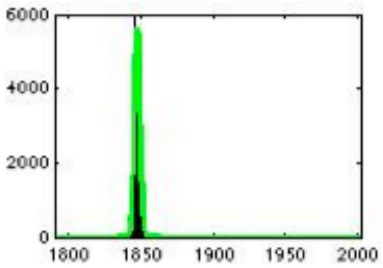
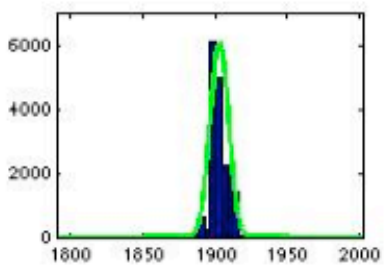
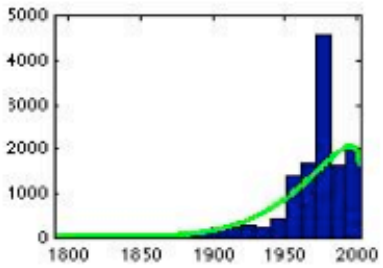
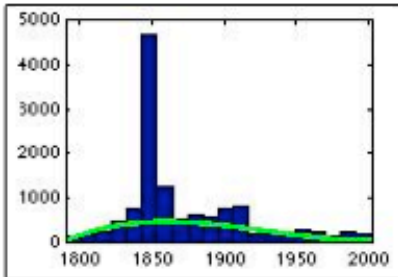
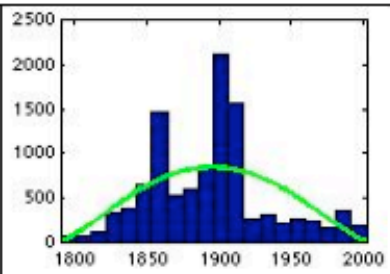
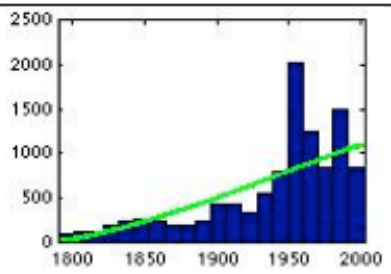
1910

Comparing

TOT

against

LDA

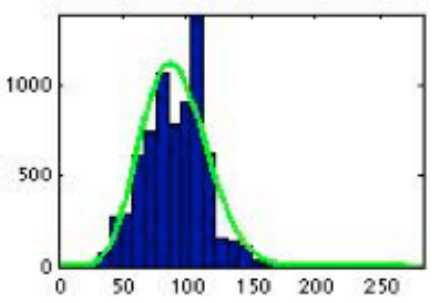
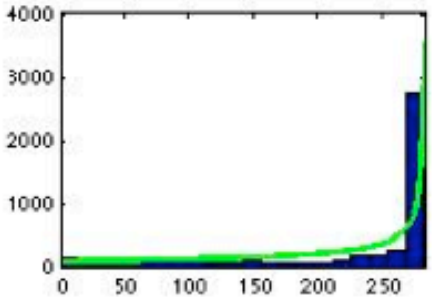
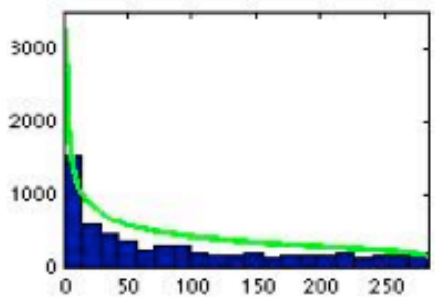
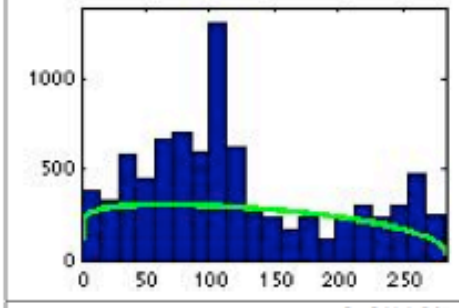
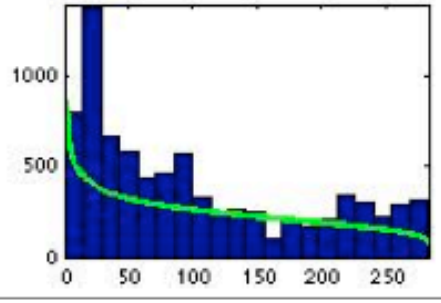
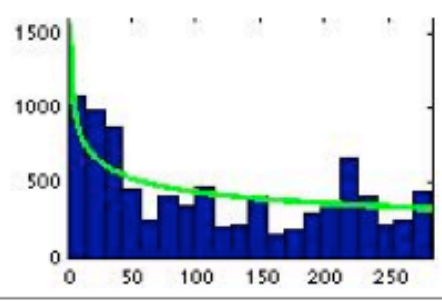
Mexican War		Panama Canal		Cold War	
					
states	0.020	government	0.029	world	0.019
mexico	0.018	united	0.021	states	0.017
government	0.017	states	0.021	security	0.017
united	0.015	islands	0.012	soviet	0.017
war	0.011	canal	0.010	united	0.015
congress	0.010	american	0.009	nuclear	0.015
country	0.009	cuba	0.008	peace	0.014
texas	0.009	made	0.007	nations	0.011
made	0.007	general	0.007	international	0.010
great	0.006	war	0.007	america	0.010
					
mexico	0.067	government	0.056	defense	0.056
government	0.023	american	0.027	military	0.038
mexican	0.021	central	0.025	forces	0.033
texas	0.021	canal	0.023	security	0.030
territory	0.017	republic	0.022	strength	0.024
part	0.016	america	0.022	nuclear	0.019
republic	0.013	pacific	0.018	weapons	0.017
military	0.011	panama	0.018	arms	0.013
state	0.010	nicaragua	0.014	maintain	0.012
make	0.009	isthmus	0.011	strong	0.011

TOT

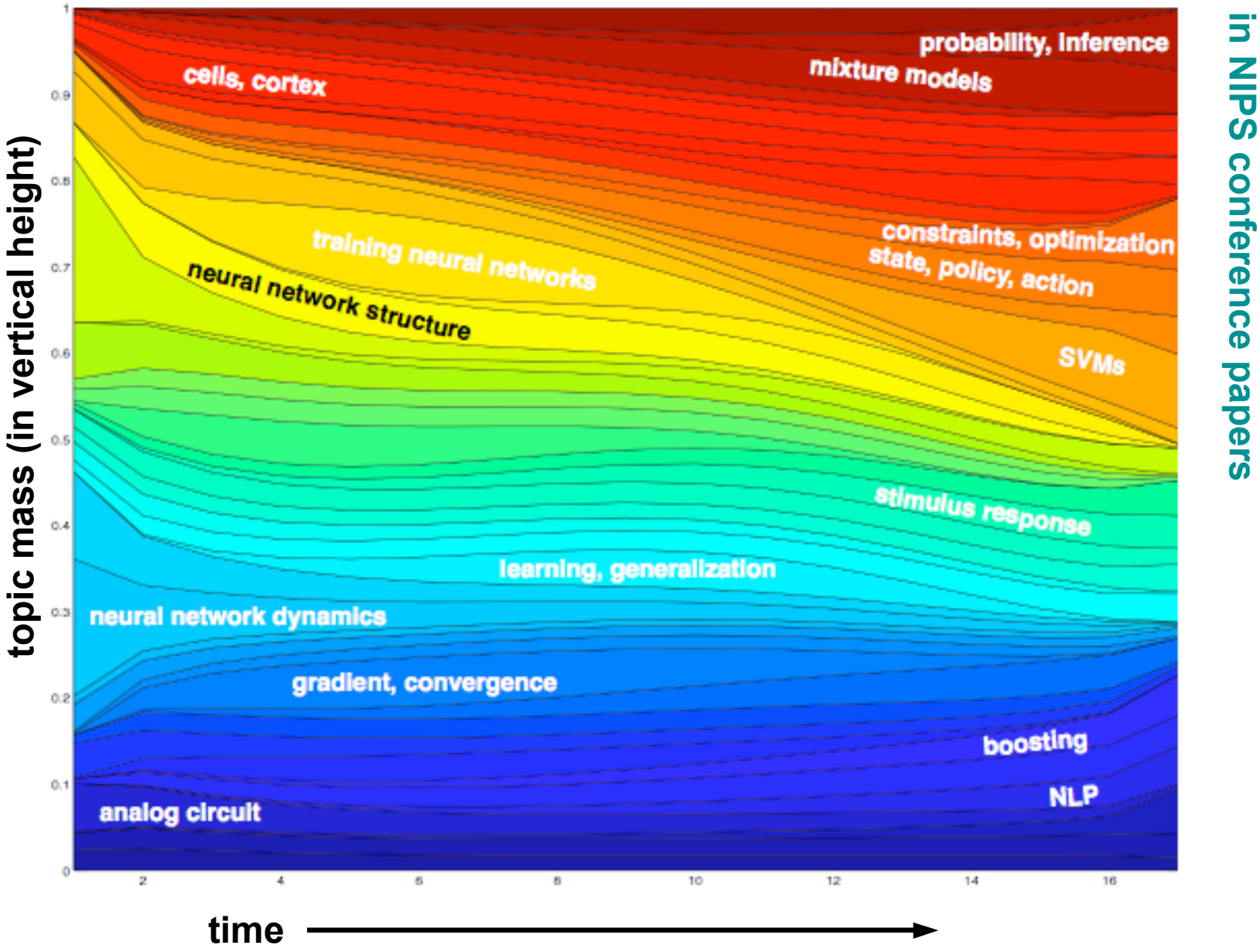
versus

LDA

on my email

Faculty Recruiting		ART Paper		MALLET	
					
cs	0.03572	xuerui	0.02113	code	0.05668
april	0.02724	data	0.01814	files	0.04212
faculty	0.02341	word	0.01601	mallet	0.04073
david	0.02012	research	0.01408	java	0.03085
lunch	0.01766	topic	0.01366	file	0.02947
schedule	0.01656	model	0.01238	al	0.02479
candidate	0.01560	andres	0.01238	directory	0.02080
talk	0.01355	sample	0.01152	version	0.01664
bruce	0.01273	enron	0.01067	pdf	0.01421
visit	0.01232	dataset	0.00960	bug	0.01352
					
cs	0.05137	email	0.09991	code	0.05947
david	0.04592	ron	0.04536	mallet	0.03922
bruce	0.02734	messages	0.04095	version	0.03772
lunch	0.02710	data	0.03408	file	0.03702
manmatha	0.02391	calo	0.03236	files	0.02534
andrew	0.02332	message	0.03053	java	0.02522
faculty	0.01764	enron	0.03028	cvs	0.02511
april	0.01740	project	0.02415	directory	0.01978
shlomo	0.01657	send	0.02023	add	0.01932
al	0.01621	part	0.01680	checked	0.01481

Topic Distributions Conditioned on Time



in NIPS conference papers

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*

  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

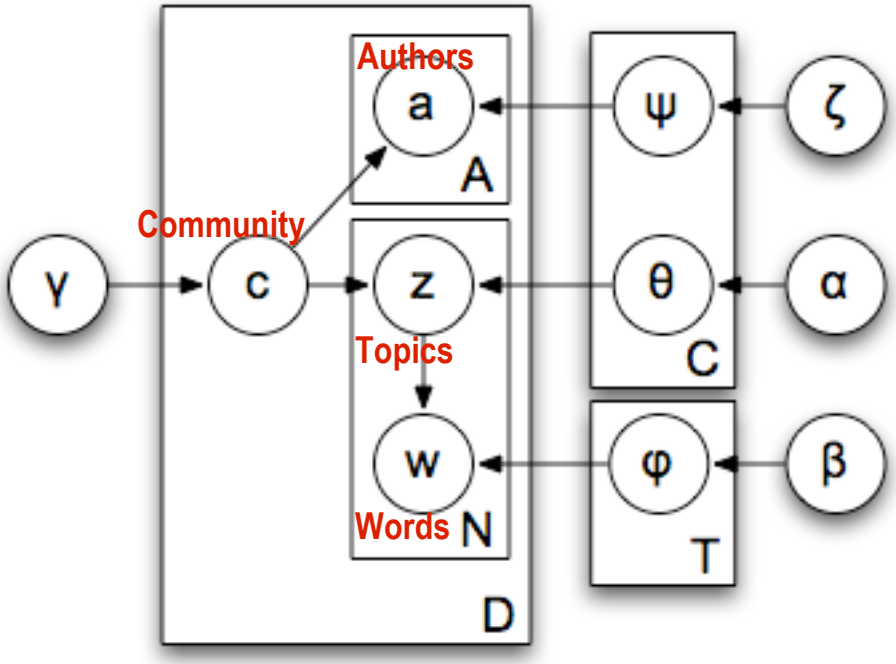
How do new links form in social networks?

1) Randomly *(Poisson graph)*
2) Pick someone popular *(Preferential attachment)*
3) Pick someone with mutual friends
   (Adamic & Adar, Liben-Nowell & Kleinberg)

4) Pick someone from one of your "communities"
   (Mimno, Wallach & McCallum 2007)

Can we find communities that help predict links?

A Community-based Generative Model for Text and Co-authorships

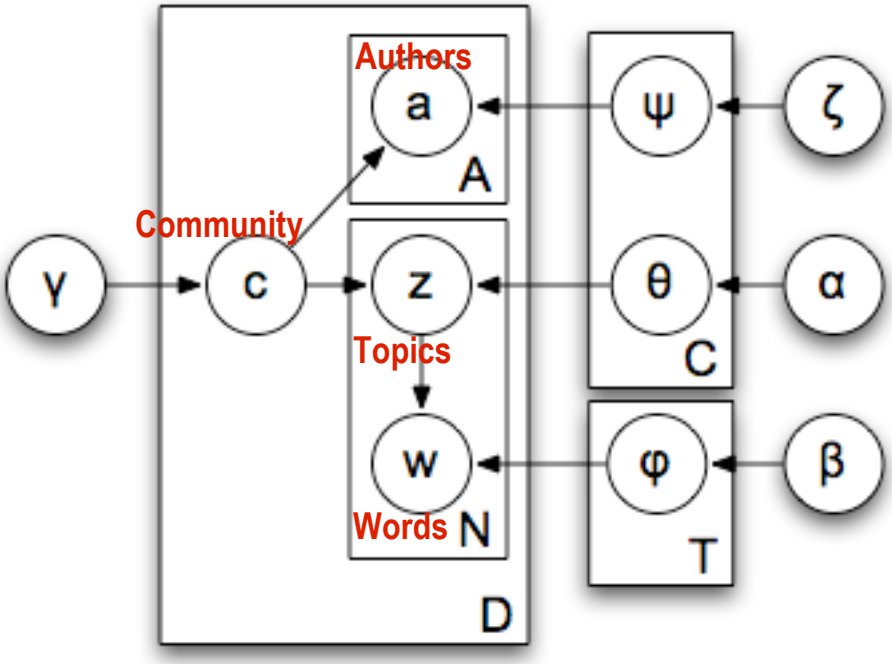


1) To generate a document, we first pick a community.

2) The community then determines the choice of authors and topics.

3) From topics, we pick words.

A Community-based Generative Model for Text and Co-authorships



Graphical Model can answer various queries!

P(author$_3$ | author$_1$, author$_2$)

P(author$_3$ | author$_1$, author$_2$, text)

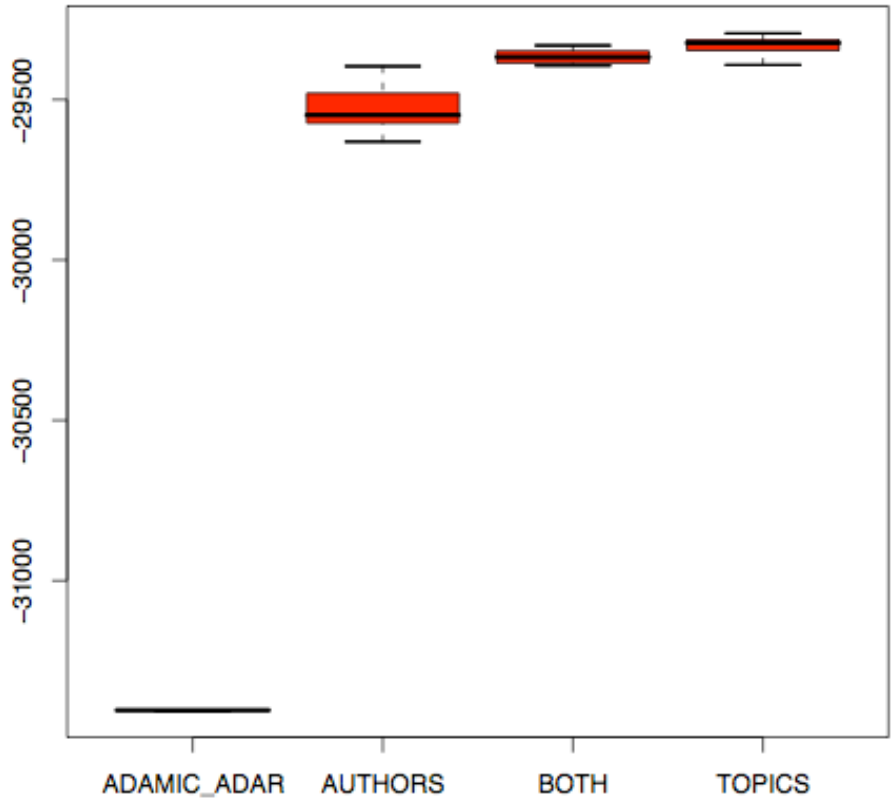
P(community | authors)

P (authors | community)

P (text | community)

P (text | authors)

Link Prediction
Probability of NIPS 2004-6 Co-authorships



(Preferential attachment is much worse, at -40,121.)

Community-Author View

Ng_A	features, feature, markov, sequence, models, conditional, label, function, set
Koller_D	number, results, paper, based, function, previous, resulting, introduction, general
Parr_R	policy, learning, action, states, function, reward, actions, optimal, mdp
Abbeel_P	control, controller, model, helicopter, system, neural, forward, learning, systems
Jordan_M	model, models, press, shows, figure, related, journal, underlying, correspond
Merzenich_M	present, effect, figure, references, important, increase, similar, addition, increased
Mel_B	learning, control, reinforcement, sutton, action, space, task, trajectory, methods

Jordan_M	propagation, belief, tree, nodes, node, approximation, variational, networks, bound
Jaakkola_T	number, results, paper, based, function, previous, resulting, introduction, general
Saul_L	theorem, case, proof, function, assume, set, section, algorithm, bound
Bach_F_R	field, boltzmann, approximations, exact, jordan, parameters, set, step, network
Singh_S	log, models, inference, variables, model, distribution, variational, parameters, matr
Wainwright_M	problem, algorithm, optimization, methods, solution, method, problems, proposed,
Nguyen_X	clustering, spectral, graph, matrix, cut, data, clusters, eigenvectors, normalized

Community-Author-Topic View

Griffiths_T_L	words, model, word, documents, document, text, topic, distribution, mixture
Singer_Y	suffix, algorithm, feature, adaptor, space, model, kernels, strings, natural
Blei_D	learning, category, naive, definition, estimation, single, figure, applied, obtain
Goldwater_S	set, labels, analysis, adclus, pmm, function, evaluation, problem, alphabet
Jordan_M	number, results, paper, based, function, previous, resulting, introduction, general
Johnson_M	prior, posterior, distribution, bayesian, likelihood, data, models, probability, model
Campbell_W	target, task, visual, figure, contrast, attention, search, orientation, discrimination
Jordan_M	propagation, belief, tree, nodes, node, approximation, variational, networks, bound
Willsky_A	field, boltzmann, approximations, exact, jordan, parameters, set, step, network
Jaakkola_T	log, models, inference, variables, model, distribution, variational, parameters, matr
Saul_L	network, variables, node, inference, distribution, nodes, algorithm, message, tree
Wiegerinck_W	number, results, paper, based, function, previous, resulting, introduction, general
Kappen_H	theorem, case, proof, function, assume, set, section, algorithm, bound
Wainwright_M	mixture, data, gaussian, density, likelihood, parameters, distribution, model, functi
Kawato_M	control, motor, learning, arm, model, movement, feedback, movements, hand
Jordan_M	eye, vor, visual, desired, field, controller, force, cerebellum, vestibular
Barto_A	neural, data, activity, figure, firing, movement, motor, speech, dynamics
Vatikiotis	present, effect, figure, references, important, increase, similar, addition, increased

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

Want a "topic model" with the advantages of Conditional Random Fields

- Use arbitrary, overlapping features of the input.
- Undirected graphical model, so we don't have to think about avoiding cycles.
- Integrate naturally with our other CRF components.
- Train "discriminatively"
- Natural semi-supervised training

What does this mean?
Topic models are unsupervised!

"Multi-Conditional Mixtures"

Latent Variable Models fit by Multi-way Conditional Probability

[McCallum, Wang, Pal, 2005],
[McCallum, Pal, Wang, 2006]

- For clustering structured data,
  ala Latent Dirichlet Allocation & its successors

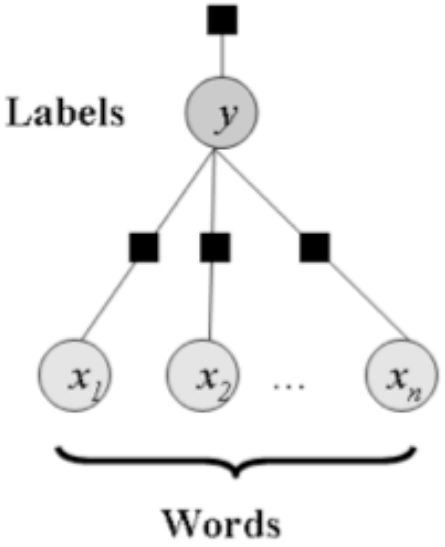
- But an undirected model,
  like the Harmonium [Welling, Rosen-Zvi, Hinton, 2005]

- But trained by a "multi-conditional" objective:
  $O = P(A|B,C)\ P(B|A,C)\ P(C|A,B)$
  e.g. A,B,C are different modalities

"Multi-Conditional Learning" (Regularization)

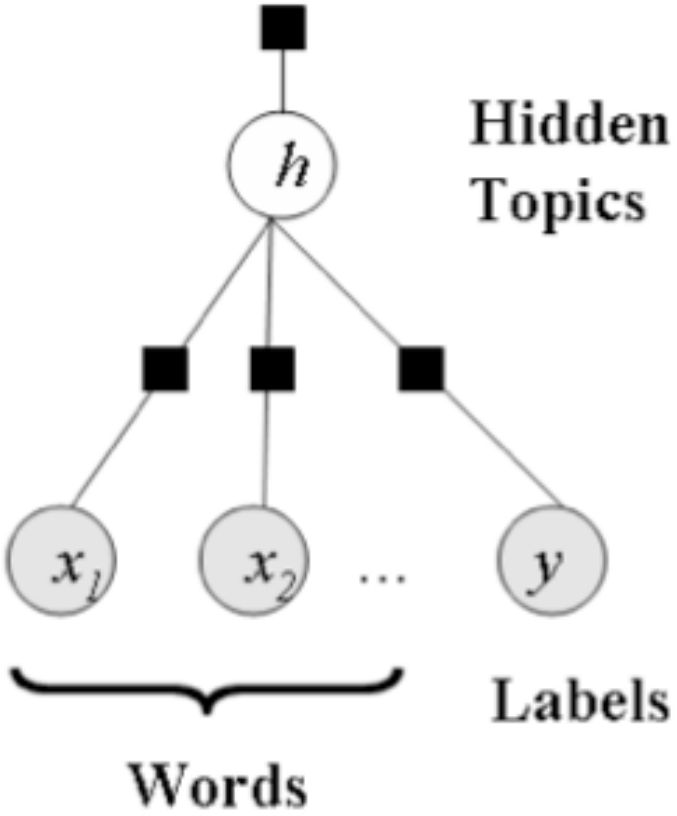
[McCallum, Pal, Wang, 2006]

$$p(y|x)p(x|y)^{\alpha}$$



Data	Naive Bayes	MaxEnt	MCL
news	85.3 (0.61)	82.9 (0.82)	**85.9 (0.89)**
news (2000)	76.4 (0.88)	77.4 (0.81)	**77.7 (0.48)**
comp	**85.1 (1.78)**	83.7 (0.68)	83.4 (0.94)
comp (2000)	81.8 (1.36)	82.2 (0.75)	**84.0 (1.05)**
talk	**84.6 (1.02)**	82.3 (1.43)	83.7 (1.27)
talk (2000)	83.7 (2.17)	81.6 (2.27)	**84.3 (1.21)**
sector	75.6 (2.05)	**88.0** (1.13)	87.4 (0.84)
sector (2000)	73.9 (0.78)	82.0 (1.03)	**83.2 (1.56)**
tech	91.0 (1.33)	.91.8 (2.24)	**93.1 (1.69)**
tech (2000)	92.9 (2.46)	91.4 (2.03)	**94.5 (1.81)**
finan	**92.3 (2.36)**	89.2 (1.52)	91.5 (2.57)
finan (2000)	87.3 (3.31)	89.6 (1.82)	**94.6 (1.79)**
health	93.5 (4.36)	94.0 (3.74)	**95.5 (4.00)**
health (2000)	95.0 (5.00)	91.0 (3.39)	**95.5 (4.30)**
movie	78.6 (1.20)	82.6 (2.96)	**82.7 (2.50)**
movie (2000)	90.9 (1.98)	88.8 (1.96)	**94.0 (1.05)**
sraa	95.9 (0.15)	96.1 (0.23)	**96.7 (0.09)**
sraa (2000)	93.7 (0.20)	94.7 (0.13)	**95.0 (0.21)**
webkb	87.9 (2.14)	**92.4 (0.84)**	**92.4 (1.04)**
webkb (2000)	84.7 (1.20)	92.4 (1.07)	**92.7 (1.40)**

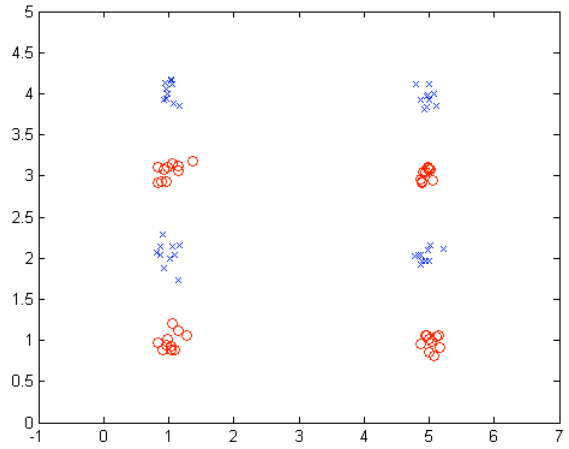
Multi-Conditional Mixtures



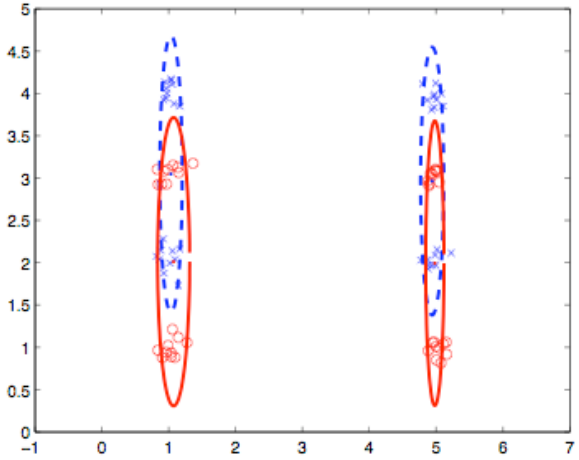
Predictive Random Fields
mixture of Gaussians on synthetic data

[McCallum, Wang, Pal, 2005]

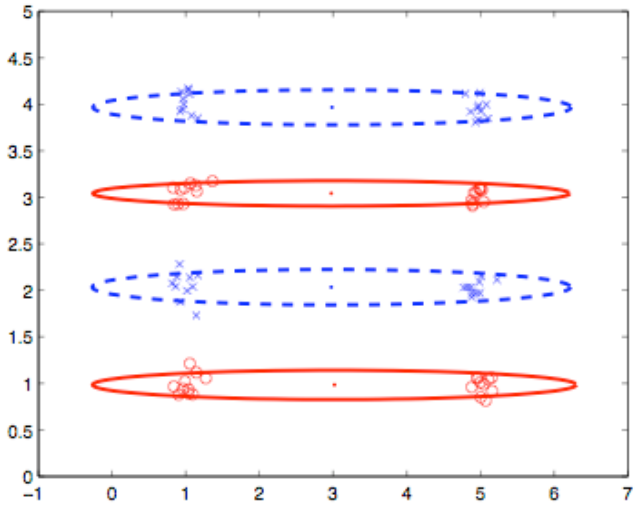
Data, classify by color



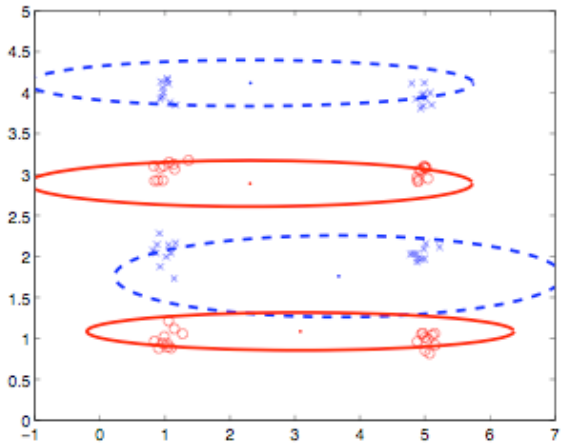
Generatively trained

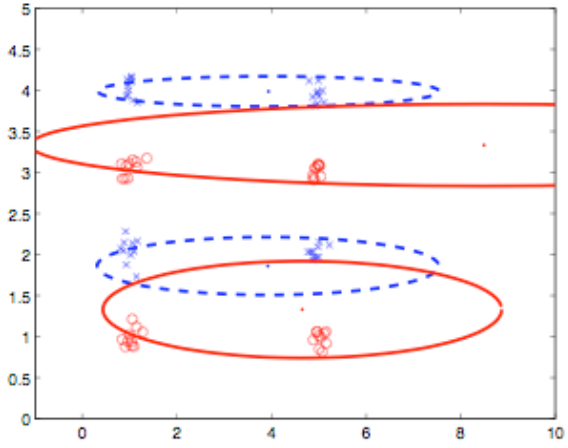


Multi-Conditional



Conditionally-trained [Jebara 1998]

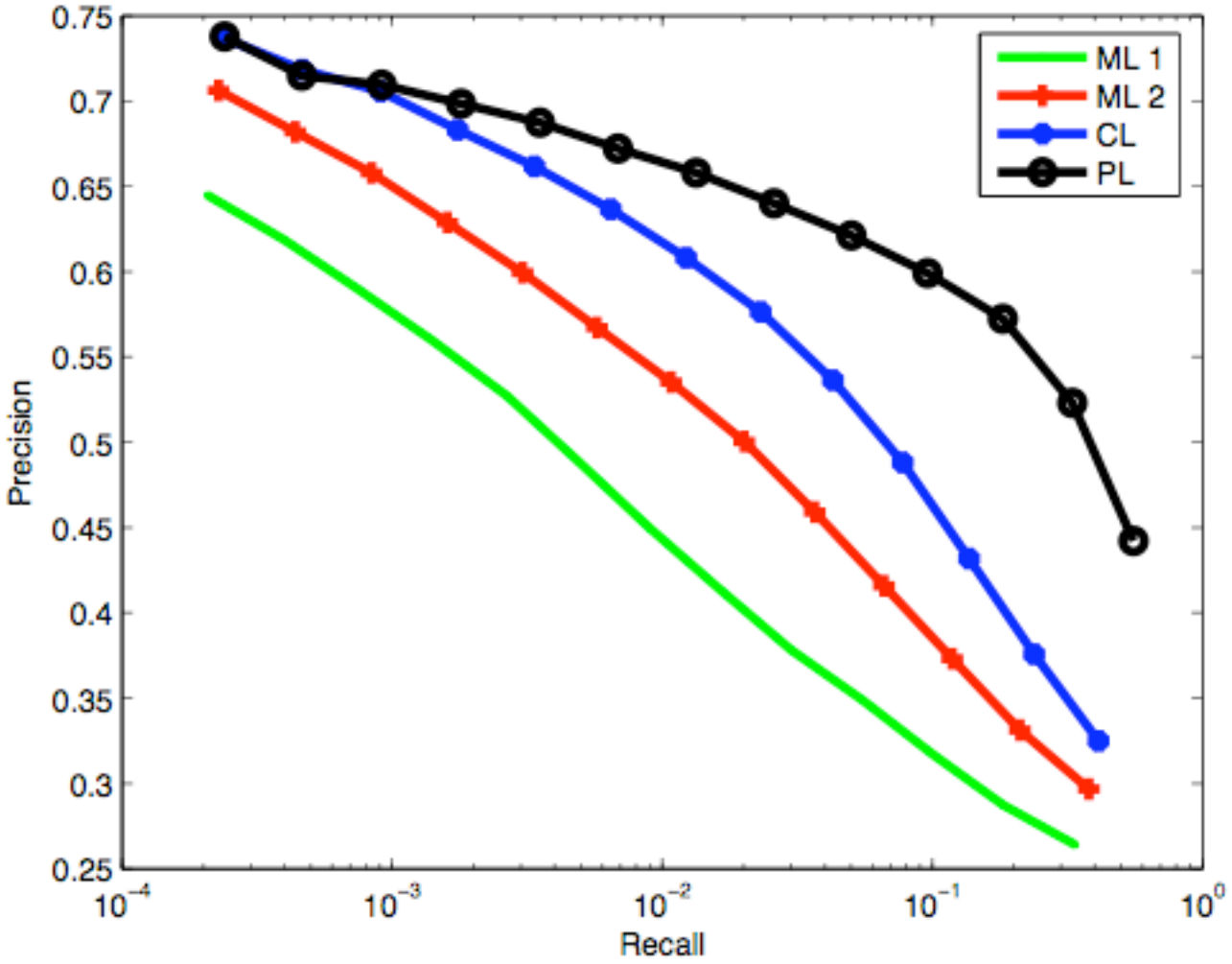




Multi-Conditional Mixtures vs. Harmoniun

on document retrieval task

[McCallum, Wang, Pal, 2005]



Multi-Conditional, multi-way conditionally trained

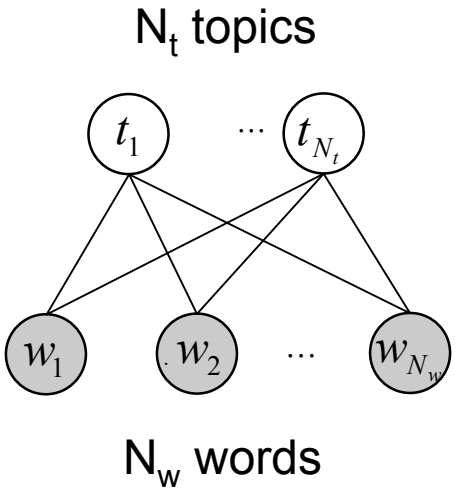
Conditionally-trained, to predict class labels

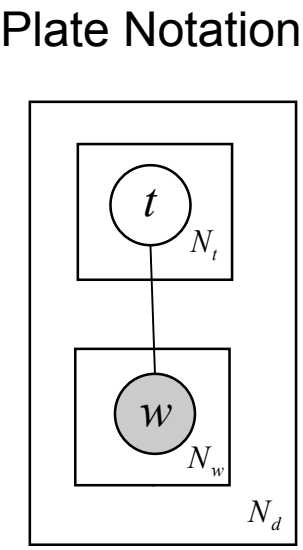
Harmonium, joint, with class labels and words

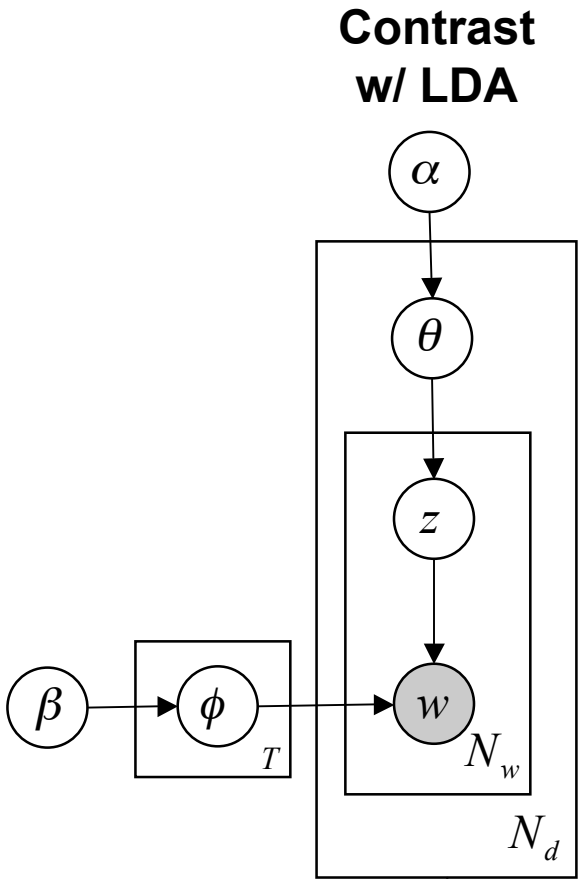
Harmonium, joint with words, no labels

Multiple Topics per Document: *"Multinomial Components Analysis" (MCA)*

- Undirected (Random Field) Topic Model
- Conditionally log-Normal Topics
- Conditionally Multinomial Words



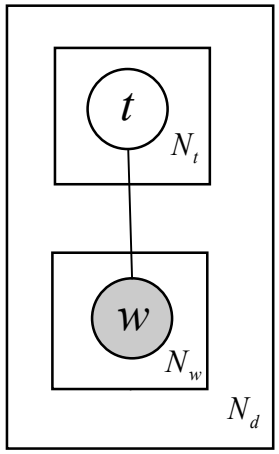




Further Contrast - MCA, PCA, RAP

- Multinomial Component Analysis (MCA)
- Principal Component Analysis (PCA)
- Rate Adapting Poisson (RAP) Model

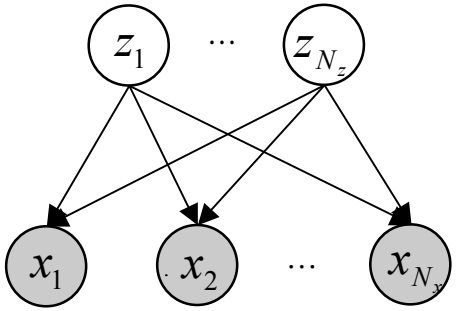
MCA



N_w draws from a discrete distribution, (words in doc)

PCA

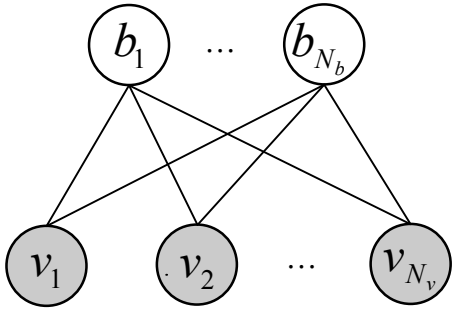
N_z unobserved, Gaussian variables



N_x observed, Gaussian variables, fixed dimension

RAP

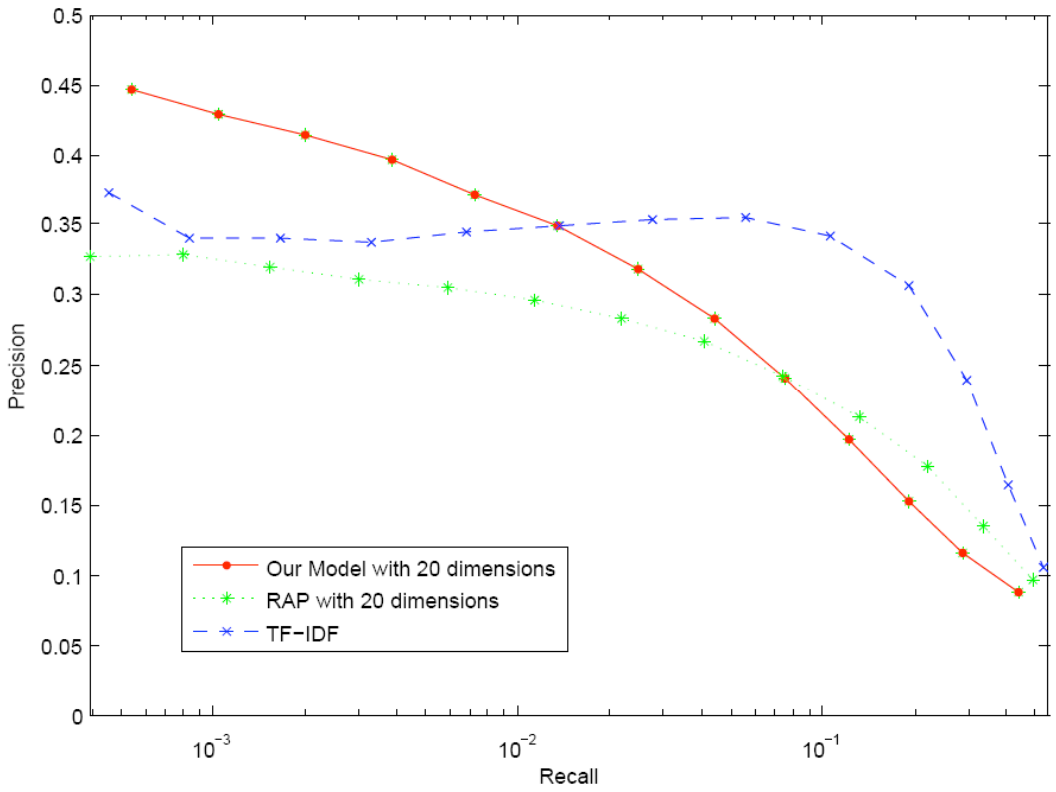
N_b binary topics



N_v Poisson counts for each word in vocabulary

Our Model (MCA) vs. TFIDF vs. RAP

MRR	Method
.45	Our Model
.37	TFIDF
.33	RAP

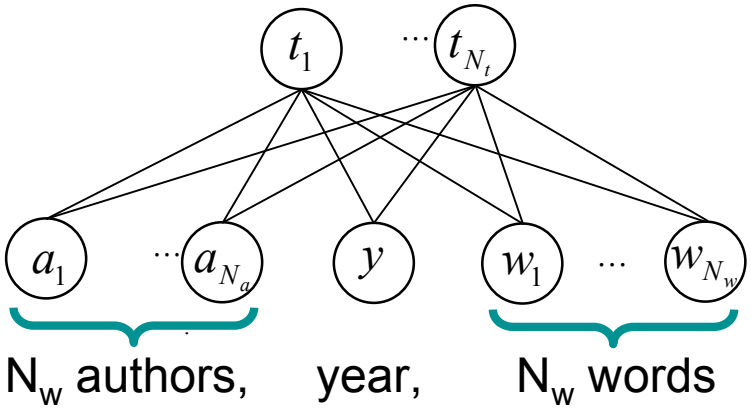


- Precision vs. Recall on 20 Newsgroups, 100 word vocabulary
- 20 dimensional hidden topic space
- Cosine Distance Comparisons (.9, .1 – Train, Test Split)
- Compared with TFIDF and Rate Adapting Poisson (RAP) Model

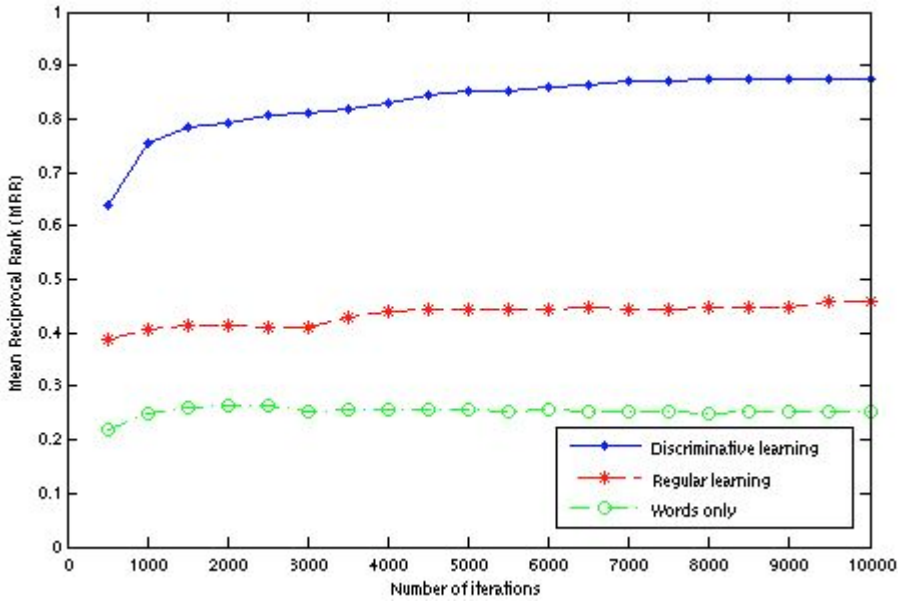
NIPS Topics

Biological Neuroscience				Reinforcement Learning				Probabilistic Methods			
cells	.439	training	-.556	learning	.318	image	-.536	data	.364	state	-.512
cell	.361	networks	-.500	policy	.266	data	-.444	model	.307	time	-.454
firing	.360	error	-.472	reinforcement	.252	images	-.431	mixture	.271	neuron	-.449
cortex	.357	network	-.470	control	.239	recognition	-.345	gaussian	.260	neural	-.429
cortical	.355	speech	-.465	state	.234	feature	-.315	likelihood	.225	system	-.422
stimulus	.327	neural	-.461	action	.233	object	-.271	image	.221	control	-.405
spike	.314	classifier	-.436	actions	.158	visual	-.270	distribution	.217	neurons	-.373
synaptic	.310	class	-.412	weight	.153	features	-.263	bayesian	.213	analog	-.363
synapses	.275	word	-.410	states	.151	gaussian	-.241	images	.204	network	-.359
motion	.268	state	-.407	controller	.150	classification	-.233	em	.189	circuit	-.335
orientation	.262	recognition	-.406	optimal	.125	mixture	-.227	density	.183	action	-.334
excitatory	.255	classifiers	-.386	weights	.121	models	-.217	models	.182	synaptic	-.317
visual	.253	classification	-.370	error	.117	model	-.211	posterior	.163	chip	-.316
inhibitory	.243	set	-.359	time	.115	likelihood	-.190	prior	.148	networks	-.287
response	.243	hmm	-.354	neuron	.105	set	-.189	regression	.146	states	-.285
stimuli	.240	algorithm	-.344	sutton	.102	orientation	-.184	kernel	.144	memory	-.279
spatial	.238	hidden	-.342	gradient	.101	classifier	-.180	log	.135	recurrent	-.263
direction	.233	test	-.337	recurrent	.101	face	-.179	classification	.134	current	-.263
membrane	.231	mixture	-.334	agent	.096	class	-.171	class	.133	policy	-.259
eye	.229	data	-.333	learn	.096	test	-.169	parameters	.124	reinforcement	-.256

Richer Model with Multiple Modalities



Predicting NIPS Authors



MRR	Method
.88	Discriminative
.46	Joint
.25	Joint, Words only

- Comparing Models, Mean Reciprocal Rank (MRR)
- Cosine Distance Comparisons (.9, .1 – Train, Test Split)

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

Social Networks in Research Literature

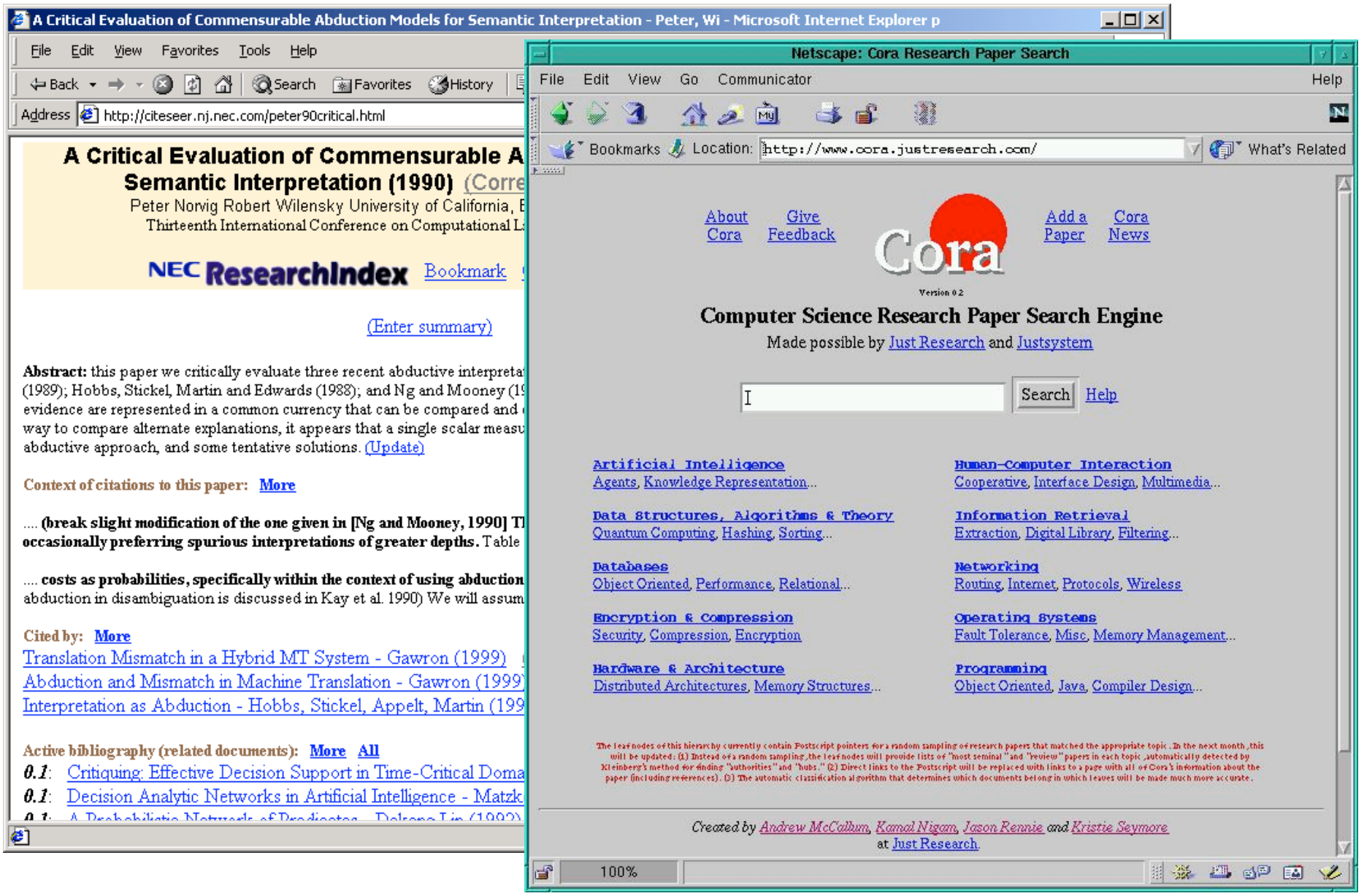
- Better understand structure of our own research area.
- Structure helps us learn a new field.
- Aid collaboration
- Map how ideas travel through social networks of researchers.

- Aids for hiring and finding reviewers!
- Measure impact of papers or people.

Our Data

- Over 1.6 million research papers, gathered as part of *Rexa.info* portal.
- Cross linked references / citations.

Previous Systems



Google Scholar BETA

"conditional random fields" Search

Scholar

Results **1** - **10** of about **154** for **"conditional random fields"**. (**0.09** seconds)

[PDF] **Conditional random fields**: Probabilistic models for segmenting and labeling sequence data
J Lafferty, A McCallum, F Pereira - View as HTML - Cited by 117
Page 1. **Conditional Random Fields**: Probabilistic Models. for Segmenting and Labeling Sequence Data. John Lafferty ¡. LAFFERTY @ CS . CMU . EDU. Andrew McCallum **...**
Proc. 18th International Conf. on Machine Learning, 2001 - aladdin.cs.cmu.edu - cis.upenn.edu - nlp.cs.nyu.edu - portal.acm.org - all 5 versions »

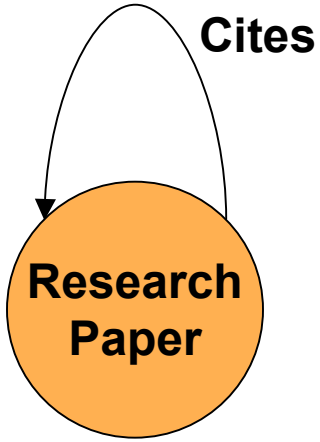
[PDF] Shallow parsing with **conditional random fields**
F Sha, F Pereira - View as HTML - Cited by 34
Page 1. Shallow Parsing with **Conditional Random Fields**. Fei Sha and Fernando Pereira Department of Computer and Information Science **...**
Proceedings of Human Language Technology, NAACL, 2003 - ldc.upenn.edu - acl.eldoc.ub.rug.nl - acl.ldc.upenn.edu - tangra.si.umich.edu - all 8 versions »

[PDF] Efficiently inducing features of **conditional random fields**
A McCallum - View as HTML - Cited by 16
Page 1. Efficiently Inducing Features of **Conditional Random Fields**. Andrew McCallum Computer Science Department University of Massachusetts **...**
Nineteenth Conference on Uncertainty in Artificial …, 2003 - ciir.cs.umass.edu - cs.umass.edu - cs.umass.edu

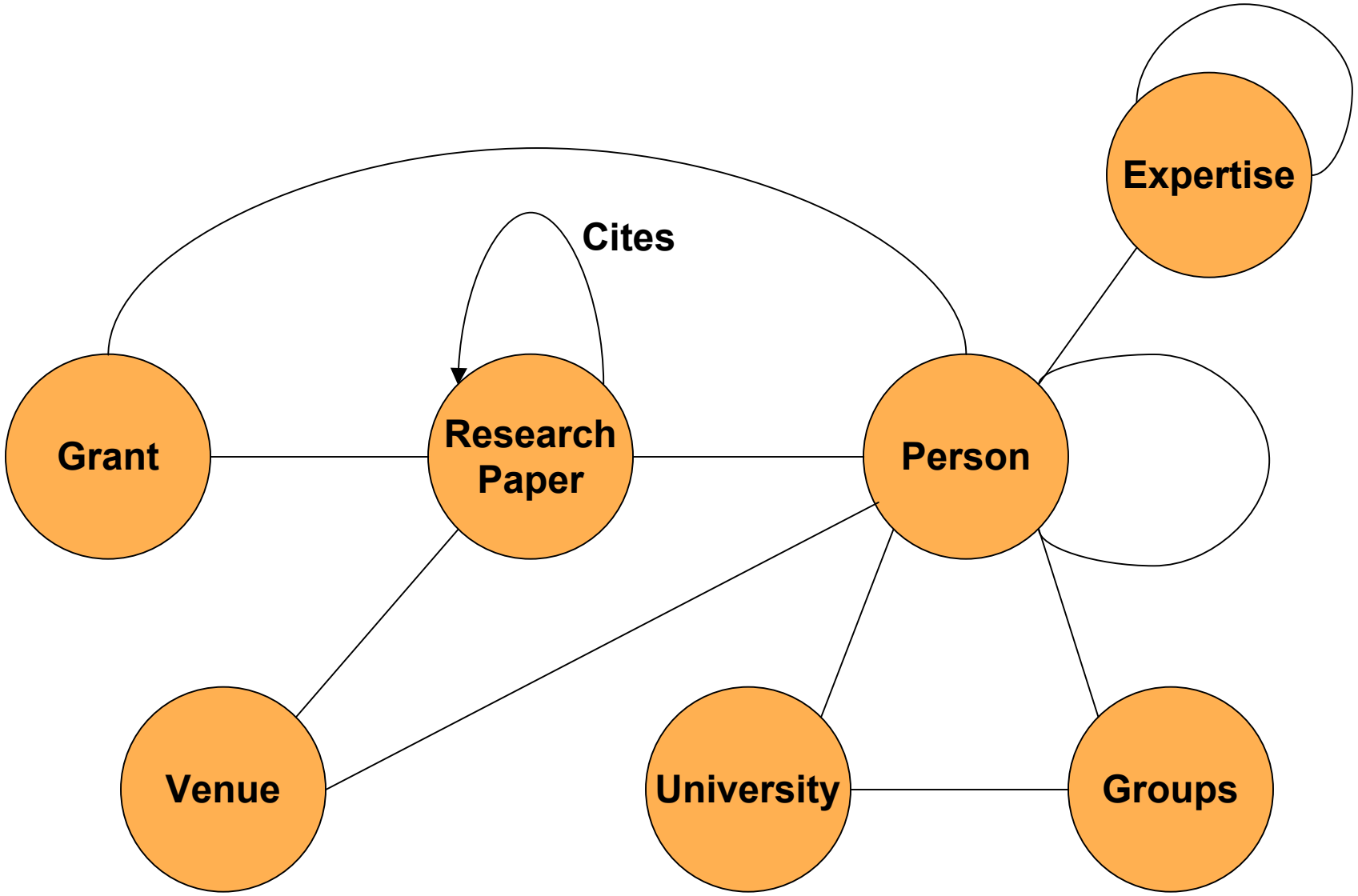
[PDF] Table extraction using **conditional random fields**
D Pinto, A McCallum, X Wei, WB Croft - Cited by 15
Page 1. Table Extraction Using **Conditional Random Fields**. David Pinto, Andrew McCallum, Xing Wei, W. Bruce Croft Center for Intelligent **...**
SIGIR'03, 2003 - portal.acm.org - cs.umass.edu - cs.umass.edu - ciir.cs.umass.edu - all 5 versions »

[PDF] Early Results for Named Entity Recognition with **Conditional Random Fields**, Feature Induction and Web …
A McCallum, W Li - View as HTML - Cited by 9
Page 1. Early Results for Named Entity Recognition with **Conditional Random Fields**, Feature Induction and Web-Enhanced Lexicons. Andrew **...**

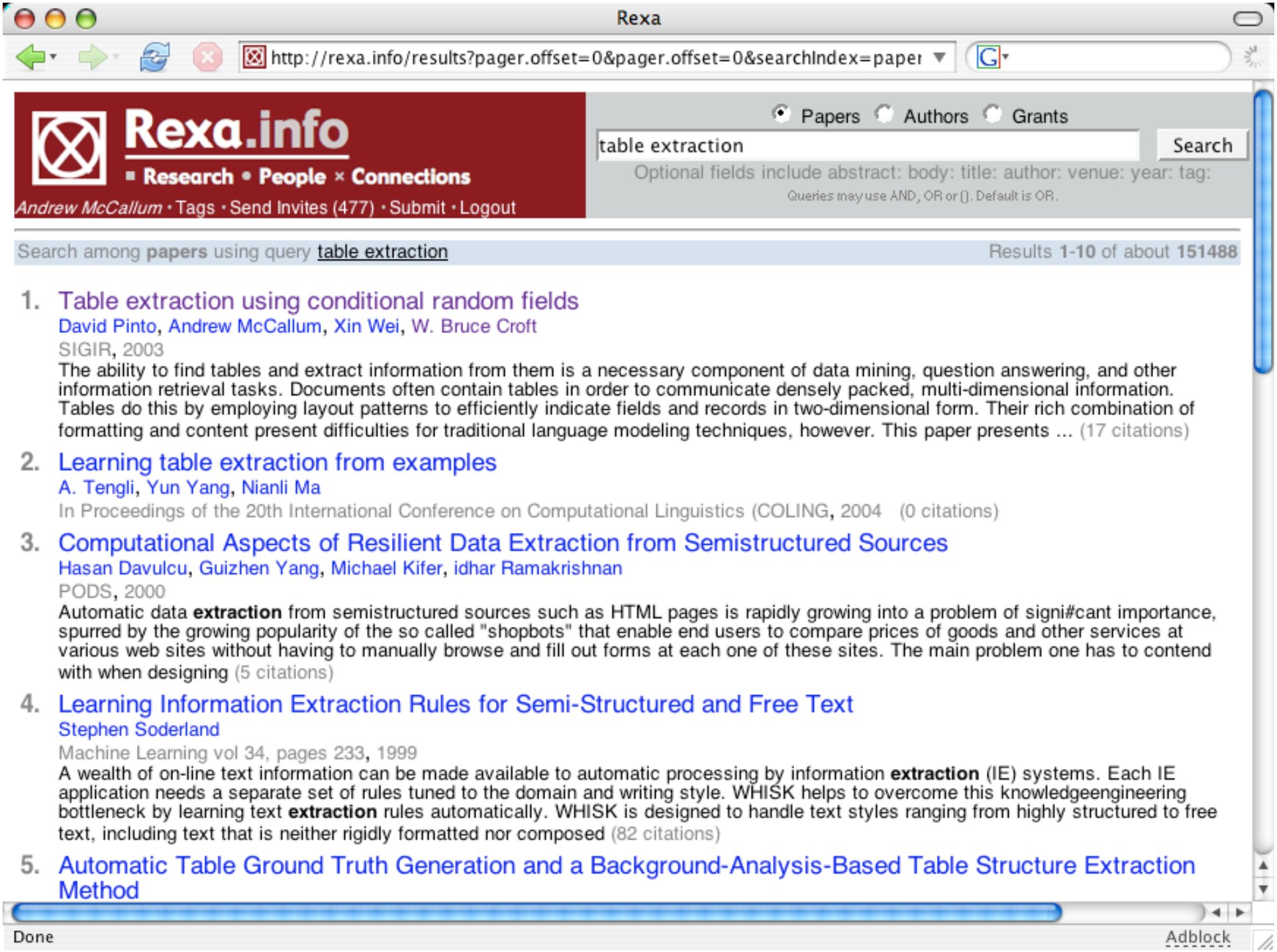
Previous Systems

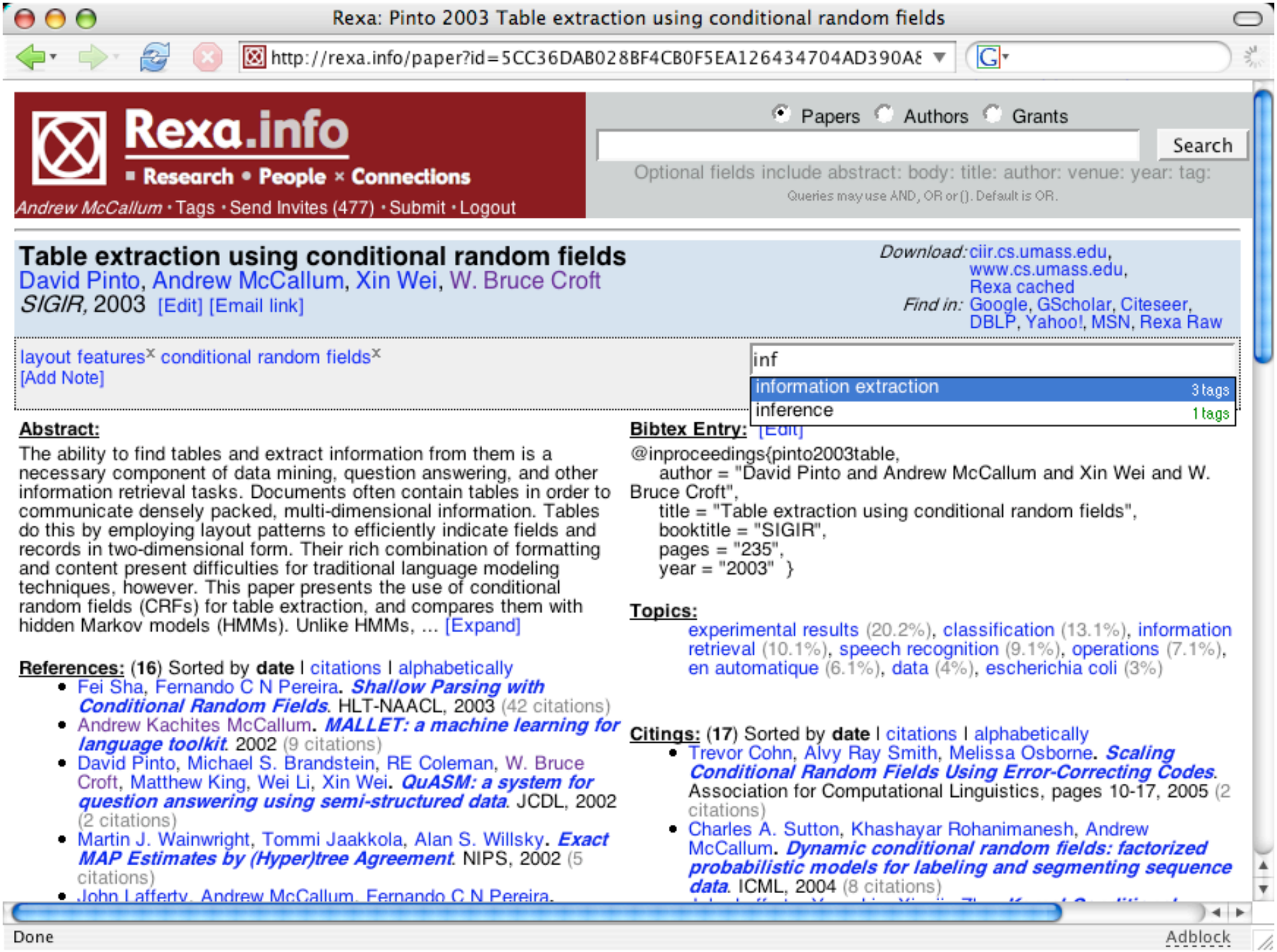


More Entities and Relations




Rexa Search Engine
http://rexa.info/
Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout
Papers Authors Grants
Search
Advanced Search
Help
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
About Statistics All Tags
Sample Queries: abstract:"reinforcement learning", author:towsley, "conditional random fields".
Search and analysis on 379,011 full text papers, 7,050,439 unique paper references & 879,678 authors.
Sponsored by :
NSF
Our work in automated information extraction and co-reference is far from finished.
Please excuse the inaccuracies and missing data while we continue our work in progress.
Version 1.0 © 2006 Created by: IESL, Department of Computer Science, University of Massachusetts
About Rexa • Help • Privacy Policy • Robot • Send feedback to rexa-discuss@cs.umass.edu.
Done
Adblock


Rexa
http://rexa.info/results?pager.offset=0&pager.offset=0&searchIndex=paper
Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout
Papers Authors Grants
table extraction
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
Search among papers using query table extraction
Results 1-10 of about 151488
1. Table extraction using conditional random fields
David Pinto, Andrew McCallum, Xin Wei, W. Bruce Croft
SIGIR, 2003
The ability to find tables and extract information from them is a necessary component of data mining, question answering, and other information retrieval tasks. Documents often contain tables in order to communicate densely packed, multi-dimensional information. Tables do this by employing layout patterns to efficiently indicate fields and records in two-dimensional form. Their rich combination of formatting and content present difficulties for traditional language modeling techniques, however. This paper presents ... (17 citations)
2. Learning table extraction from examples
A. Tengli, Yun Yang, Nianli Ma
In Proceedings of the 20th International Conference on Computational Linguistics (COLING, 2004 (0 citations)
3. Computational Aspects of Resilient Data Extraction from Semistructured Sources
Hasan Davulcu, Guizhen Yang, Michael Kifer, idhar Ramakrishnan
PODS, 2000
Automatic data extraction from semistructured sources such as HTML pages is rapidly growing into a problem of signi#cant importance, spurred by the growing popularity of the so called "shopbots" that enable end users to compare prices of goods and other services at various web sites without having to manually browse and fill out forms at each one of these sites. The main problem one has to contend with when designing (5 citations)
4. Learning Information Extraction Rules for Semi-Structured and Free Text
Stephen Soderland
Machine Learning vol 34, pages 233, 1999
A wealth of on-line text information can be made available to automatic processing by information extraction (IE) systems. Each IE application needs a separate set of rules tuned to the domain and writing style. WHISK helps to overcome this knowledgeengineering bottleneck by learning text extraction rules automatically. WHISK is designed to handle text styles ranging from highly structured to free text, including text that is neither rigidly formatted nor composed (82 citations)
5. Automatic Table Ground Truth Generation and a Background-Analysis-Based Table Structure Extraction Method
Done
Adblock


Rexa: Pinto 2003 Table extraction using conditional random fields
http://rexa.info/paper?id=5CC36DAB028BF4CB0F5EA126434704AD390A8
Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout
Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
Table extraction using conditional random fields
David Pinto, Andrew McCallum, Xin Wei, W. Bruce Croft
SIGIR, 2003 [Edit] [Email link]
Download: ciir.cs.umass.edu, www.cs.umass.edu, Rexa cached
Find in: Google, GScholar, Citeseer, DBLP, Yahoo!, MSN, Rexa Raw
layout features conditional random fields
[Add Note]
inf
information extraction 3 tags
inference 1 tags
Abstract:
The ability to find tables and extract information from them is a necessary component of data mining, question answering, and other information retrieval tasks. Documents often contain tables in order to communicate densely packed, multi-dimensional information. Tables do this by employing layout patterns to efficiently indicate fields and records in two-dimensional form. Their rich combination of formatting and content present difficulties for traditional language modeling techniques, however. This paper presents the use of conditional random fields (CRFs) for table extraction, and compares them with hidden Markov models (HMMs). Unlike HMMs, ... [Expand]
Bibtex Entry: [Edit]
@inproceedings{pinto2003table,
author = "David Pinto and Andrew McCallum and Xin Wei and W. Bruce Croft",
title = "Table extraction using conditional random fields",
booktitle = "SIGIR",
pages = "235",
year = "2003" }
Topics:
experimental results (20.2%), classification (13.1%), information retrieval (10.1%), speech recognition (9.1%), operations (7.1%), en automatique (6.1%), data (4%), escherichia coli (3%)
References: (16) Sorted by date | citations | alphabetically
Fei Sha, Fernando C N Pereira. Shallow Parsing with Conditional Random Fields. HLT-NAACL, 2003 (42 citations)
Andrew Kachites McCallum. MALLET: a machine learning for language toolkit. 2002 (9 citations)
David Pinto, Michael S. Brandstein, RE Coleman, W. Bruce Croft, Matthew King, Wei Li, Xin Wei. QuASM: a system for question answering using semi-structured data. JCDL, 2002 (2 citations)
Martin J. Wainwright, Tommi Jaakkola, Alan S. Willsky. Exact MAP Estimates by (Hyper)tree Agreement. NIPS, 2002 (5 citations)
John Lafferty, Andrew McCallum, Fernando C N Pereira,
Citings: (17) Sorted by date | citations | alphabetically
Trevor Cohn, Alvy Ray Smith, Melissa Osborne. Scaling Conditional Random Fields Using Error-Correcting Codes. Association for Computational Linguistics, pages 10-17, 2005 (2 citations)
Charles A. Sutton, Khashayar Rohanimanesh, Andrew McCallum. Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data. ICML, 2004 (8 citations)
Done
Adblock

Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout

Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.

W. Bruce Croft [Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

Publications: (*1 to 40* of 233) *(total 1436 citations)*
Sorted by **date** | citations

2004
- Donald Metzler, W. Bruce Croft. ***Combining the language model and inference network approaches to retrieval***. Inf. Process. Manage. vol 40, pages 735, 2004 (1 citation)
- Xiaoyong Liu, W. Bruce Croft. ***Cluster-based retrieval using language models***. SIGIR, 2004 (0 citations)
- Andrés Corrada-Emmanuel, W. Bruce Croft. ***Answer models for question answering passage retrieval***. SIGIR, 2004 (0 citations)
- Chirag Shah, W. Bruce Croft. ***Evaluating high accuracy retrieval techniques***. SIGIR, 2004 (1 citation)
- Haizheng Zhang, W. Bruce Croft, Brian N. Levine, Victor R. Lesser. ***A Multi-Agent Approach for Peer-to-Peer Based Information Retrieval System***. AAMAS, 2004 (1 citation)
- Donald Metzler, Victor Lavrenko, W. Bruce Croft. ***Formal multiple-bernoulli models for language modeling***. SIGIR, 2004 (0 citations)
- Stephen Cronen-Townsend, Yu Zhou, W. Bruce Croft. ***A framework for selective query expansion***. CIKM, 2004 (0 citations)

2003
- W. Bruce Croft. ***Language Models for Information Retrieval***.

Co-authors | Cited authors | Citing authors: (*1 to 40* of 257)
Sorted by **date** | number | name

- **Victor Lavrenko,** 2004 2003 2002 2002 2001 2001 ???? ????
- **Stephen Cronen-Townsend,** 2004 2002 2001 ????
- **Donald Metzler,** 2004 2004 2003
- **Xiaoyong Liu,** 2004 2002
- **Andrés Corrada-Emmanuel,** 2004
- **Victor R. Lesser,** 2004
- **Brian N. Levine,** 2004
- **Chirag Shah,** 2004
- **Haizheng Zhang,** 2004
- **Yu Zhou,** 2004
- **James P. Callan,** 2003 2001 1997 1996 1996 1995 1995 1995 1995 1995 1994 1994 1994 1994 1994 1993 1993 1993 1992 1992 ???? ???? ????
- **Howard R. Turtle,** 2003 1999 1997 1996 1993 1992

Done
Adblock

Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info
Research • People × Connections

Andrew McCallum • Tags • Send Invites (477) • Submit • Logout

Papers Authors Grants

Search

Optional fields include abstract: body: title: author: venue: year: tag:

Queries may use AND, OR or (). Default is OR.

W. Bruce Croft

[Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

Publications: (*1 to 40* of 233) *(total 1436 citations)*
Sorted by date | **citations**

~ 90
- James P. Callan, Zhihong Lu, W. Bruce Croft. ***Searching Distributed Collections with Inference Networks***. SIGIR, 1995 (84 citations)
- James P. Callan, W. Bruce Croft, Stephen M. Harding. ***The INQUERY Retrieval System***. DEXA, 1992 (80 citations)

~ 80
- Jay M. Ponte, W. Bruce Croft. ***A Language Modeling Approach to Information Retrieval***. SIGIR, 1998 (77 citations)

~ 70
- Jinxi Xu, W. Bruce Croft. ***Query Expansion Using Local and Global Document Analysis***. SIGIR, 1996 (63 citations)
- Nicholas J. Belkin, W. Bruce Croft. ***Information Filtering and Information Retrieval: Two Sides of the Same Coin***. Commun. ACM vol 35, pages 29, 1992 (63 citations)

~ 50
- Howard R. Turtle, W. Bruce Croft. ***Evaluation of an Inference Network-Based Retrieval Model***. ACM Trans. Inf. Syst. vol 9, pages 187, 1991 (48 citations)

~ 40
- Isidro Laso Ballesteros, W. Bruce Croft. ***Phrasal Translation and Query Expansion Techniques for Cross-language Information Retrieval***. SIGIR, 1997 (39 citations)
- Isidro Laso Ballesteros, W. Bruce Croft. ***Resolving Ambiguity for Cross-Language Retrieval***. SIGIR, 1998 (36 citations)

Co-authors | Cited authors | Citing authors: (*1 to 40* of 257)
Sorted by **date** | number | name

- **Victor Lavrenko,** 2004 2003 2002 2002 2001 2001 ???? ????
- **Stephen Cronen-Townsend,** 2004 2002 2001 ????
- **Donald Metzler,** 2004 2004 2003
- **Xiaoyong Liu,** 2004 2002
- **Andrés Corrada-Emmanuel,** 2004
- **Victor R. Lesser,** 2004
- **Brian N. Levine,** 2004
- **Chirag Shah,** 2004
- **Haizheng Zhang,** 2004
- **Yu Zhou,** 2004
- **James P. Callan,** 2003 2001 1997 1996 1996 1995 1995 1995 1995 1995 1994 1994 1994 1994 1994 1993 1993 1993 1992 1992 ???? ???? ????
- **Howard R. Turtle,** 2003 1999 1997 1996 1993 1992 1992 1991 1991 1991 1990 1990 1990 1989

Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout

Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.

W. Bruce Croft [Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

Publications: (*1 to 40* of 233) *(total 1436 citations)*
Sorted by **date** | citations

2004
- Donald Metzler, W. Bruce Croft. ***Combining the language model and inference network approaches to retrieval***. Inf. Process. Manage. vol 40, pages 735, 2004 (1 citation)
- Xiaoyong Liu, W. Bruce Croft. ***Cluster-based retrieval using language models***. SIGIR, 2004 (0 citations)
- Andrés Corrada-Emmanuel, W. Bruce Croft. ***Answer models for question answering passage retrieval***. SIGIR, 2004 (0 citations)
- Chirag Shah, W. Bruce Croft. ***Evaluating high accuracy retrieval techniques***. SIGIR, 2004 (1 citation)
- Haizheng Zhang, W. Bruce Croft, Brian N. Levine, Victor R. Lesser. ***A Multi-Agent Approach for Peer-to-Peer Based Information Retrieval System***. AAMAS, 2004 (1 citation)
- Donald Metzler, Victor Lavrenko, W. Bruce Croft. ***Formal multiple-bernoulli models for language modeling***. SIGIR, 2004 (0 citations)
- Stephen Cronen-Townsend, Yu Zhou, W. Bruce Croft. ***A framework for selective query expansion***. CIKM, 2004 (0 citations)

2003
- W. Bruce Croft. ***Language Models for Information Retrieval***.

Co-authors | Cited authors | Citing authors: (*1 to 40* of 257)
Sorted by date | **number** | name

- **James P. Callan,** 2003 2001 1997 1996 1996 1995 1995 1995 1995 1995 1994 1994 1994 1994 1994 1993 1993 1993 1992 1992 ???? ???? ????
- **Howard R. Turtle,** 2003 1999 1997 1996 1993 1992 1992 1991 1991 1991 1990 1990 1990 1989
- **John Broglio,** 1996 1996 1996 1995 1995 1994 1994 1994 1994 1994 1994 1993
- **Nicholas J. Belkin,** 2003 2002 1993 1992 1990 1987 1987 1987 ????
- **Victor Lavrenko,** 2004 2003 2002 2002 2001 2001 ???? ????
- **James Allan,** 2003 2003 2002 2000 1997 1996 1995
- **Jinxi Xu,** 2003 2000 1999 1998 1998 1996 1995
- **Isidro Laso Ballesteros,** 1998 1998 1998 1997 1997 1996 1996

Done
Adblock

Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info
Research • People × Connections
Andrew McCallum · Tags · Send Invites (477) · Submit · Logout

Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.

W. Bruce Croft [Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

Publications: (*1 to 40* of 233) *(total 1436 citations)*
Sorted by **date** | citations

2004
- Donald Metzler, W. Bruce Croft. ***Combining the language model and inference network approaches to retrieval***. Inf. Process. Manage. vol 40, pages 735, 2004 (1 citation)
- Xiaoyong Liu, W. Bruce Croft. ***Cluster-based retrieval using language models***. SIGIR, 2004 (0 citations)
- Andrés Corrada-Emmanuel, W. Bruce Croft. ***Answer models for question answering passage retrieval***. SIGIR, 2004 (0 citations)
- Chirag Shah, W. Bruce Croft. ***Evaluating high accuracy retrieval techniques***. SIGIR, 2004 (1 citation)
- Haizheng Zhang, W. Bruce Croft, Brian N. Levine, Victor R. Lesser. ***A Multi-Agent Approach for Peer-to-Peer Based Information Retrieval System***. AAMAS, 2004 (1 citation)
- Donald Metzler, Victor Lavrenko, W. Bruce Croft. ***Formal multiple-bernoulli models for language modeling***. SIGIR, 2004 (0 citations)
- Stephen Cronen-Townsend, Yu Zhou, W. Bruce Croft. ***A framework for selective query expansion***. CIKM, 2004 (0 citations)

2003
- W. Bruce Croft. ***Language Models for Information Retrieval***. ICDE, 2003 (0 citations)
- W. Bruce Croft, John Lafferty. ***Language Modeling for Information***

Co-authors | **Cited authors** | Citing authors: (*1 to 40* of 368)
Sorted by date | **number** | name

- **W. Bruce Croft,** *2004 2003 2002 2002 2002 2001 2000 2000 1999 1999 1998 1998 1997 1997 1997 1997 1996 1996 1996 1995 1995 1995 1995 1995 1995 1994 1994 1994 1994 1994 1993 1993 1993 1992 1992 1992 1991 1991 1991 1990 1990 1979 ???? ????*
- **James P. Callan,** 2001 1999 *1997 1995 1995 1995 1994 1994* 1994 *1994 1993 1992*
- **Ellen M. Voorhees,** 2002 2001 2000 2000 1999 1994 1993 1993 1983
- **James Allan,** 1999 1998 *1997 1995* 1993 ????
- **Howard R. Turtle,** 1994 *1992 1991 1991 1991 1990*
- **Justin Zobel,** 2001 1996 1994 1994 1992
- **John Broglio,** *1996 1995 1994 1994 1994*
- **Hector Garcia-Molina,** 1995 1994 1994 1993 ????
- **Donna Harman,** 1995 1992 1992 1991 1988

Rexa: W. Bruce Croft

http://rexa.info/author?id=DD3413947C0716FD5B95E4912C16BC8E9F72I

Rexa.info
Research • People × Connections

Andrew McCallum • Tags • Send Invites (477) • Submit • Logout

Papers Authors Grants

Search

Optional fields include abstract: body: title: author: venue: year: tag:

Queries may use AND, OR or (). Default is OR.

W. Bruce Croft

[Google][Edit Info][Send Invite][Email link]

Distinguished Professor
Department of Computer Science, University of Massachusetts
BRUCE CROFT, Amherst, MA, 01003-9264
Email: croftg@cs.umass.edu
URL: http://ciir.cs.umass.edu/personnel/croft.html

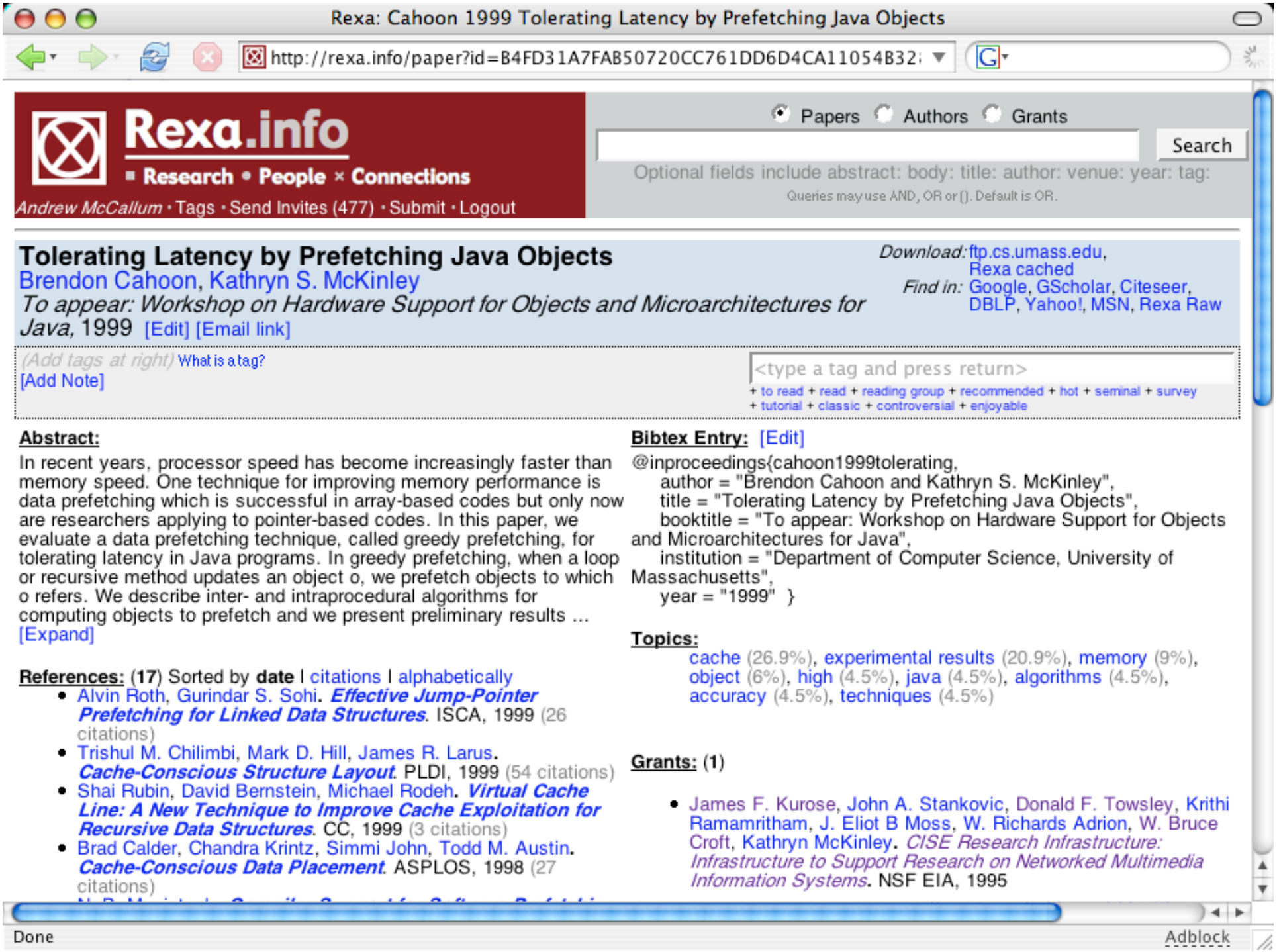
Publications: (*1 to 40* **of 233**) ***(total 1436 citations)***
Sorted by **date** | citations

2004
- Donald Metzler, W. Bruce Croft. ***Combining the language model and inference network approaches to retrieval***. Inf. Process. Manage. vol 40, pages 735, 2004 (1 citation)
- Xiaoyong Liu, W. Bruce Croft. ***Cluster-based retrieval using language models***. SIGIR, 2004 (0 citations)
- Andrés Corrada-Emmanuel, W. Bruce Croft. ***Answer models for question answering passage retrieval***. SIGIR, 2004 (0 citations)
- Chirag Shah, W. Bruce Croft. ***Evaluating high accuracy retrieval techniques***. SIGIR, 2004 (1 citation)
- Haizheng Zhang, W. Bruce Croft, Brian N. Levine, Victor R. Lesser. ***A Multi-Agent Approach for Peer-to-Peer Based Information Retrieval System***. AAMAS, 2004 (1 citation)
- Donald Metzler, Victor Lavrenko, W. Bruce Croft. ***Formal multiple-bernoulli models for language modeling***. SIGIR, 2004 (0 citations)
- Stephen Cronen-Townsend, Yu Zhou, W. Bruce Croft. ***A framework for selective query expansion***. CIKM, 2004 (0 citations)

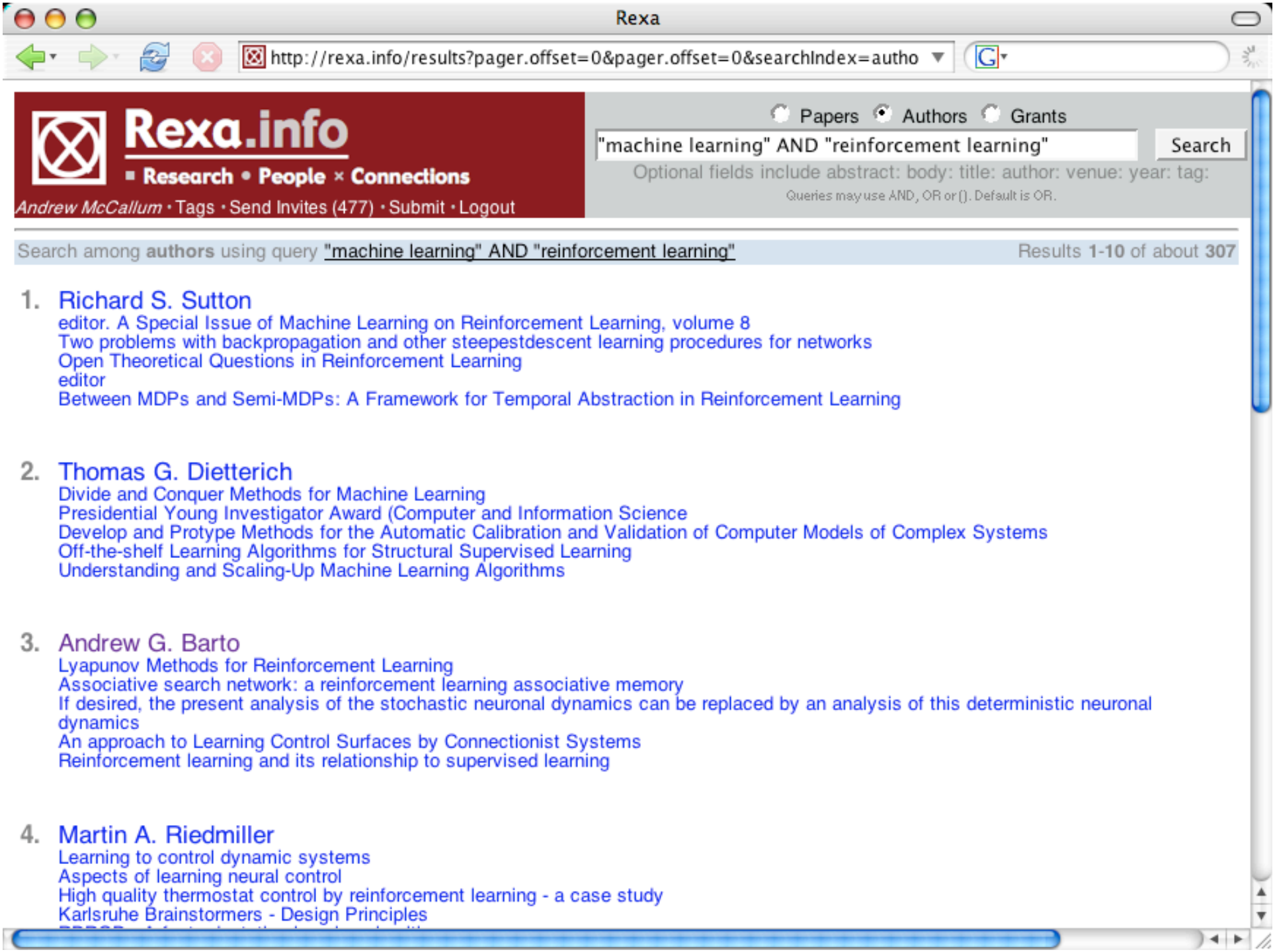
2003
- W. Bruce Croft. ***Language Models for Information Retrieval***. ICDE, 2003 (0 citations)

Co-authors | Cited authors | **Citing authors**: (*1 to 40* **of 1527**)
Sorted by date | **number** | name

- **W. Bruce Croft,** *2004 2004 2003 2002 2002 2002 2002 2001 1998 1997 1997 1997 1996 1995 1995 1994 1994 1993 1992 ???? ???? ???? ???? ???? ???? ???? ????*
- **James Allan,** 2004 2003 2002 2002 2001 2001 2000 1998 1998 1996 1996 1994 1993 ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ????
- **Douglas W. Oard,** 2003 2003 2003 2002 1999 1998 1998 1996 1996 1995 ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ???? ????
- **Victor Lavrenko,** 2004 2004 2003 *2002* 2002 *2001* 2000 1998 1996 ???? ???? ???? ???? ???? ???? *????* ???? ???? ????
- **James P. Callan,** 2004 2003 2002 2002 2001 2000 2000 1996 *1995 1994* 1994 *1994 1993 1992 ????*


Rexa: Cahoon 1999 Tolerating Latency by Prefetching Java Objects
http://rexa.info/paper?id=B4FD31A7FAB50720CC761DD6D4CA11054B32
Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout
Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
Tolerating Latency by Prefetching Java Objects
Brendon Cahoon, Kathryn S. McKinley
To appear: Workshop on Hardware Support for Objects and Microarchitectures for Java, 1999 [Edit] [Email link]
Download: ftp.cs.umass.edu, Rexa cached
Find in: Google, GScholar, Citeseer, DBLP, Yahoo!, MSN, Rexa Raw
(Add tags at right) What is a tag?
[Add Note]
<type a tag and press return>
+ to read + read + reading group + recommended + hot + seminal + survey
+ tutorial + classic + controversial + enjoyable
Abstract:
In recent years, processor speed has become increasingly faster than memory speed. One technique for improving memory performance is data prefetching which is successful in array-based codes but only now are researchers applying to pointer-based codes. In this paper, we evaluate a data prefetching technique, called greedy prefetching, for tolerating latency in Java programs. In greedy prefetching, when a loop or recursive method updates an object o, we prefetch objects to which o refers. We describe inter- and intraprocedural algorithms for computing objects to prefetch and we present preliminary results ...
[Expand]
References: (17) Sorted by date | citations | alphabetically
Alvin Roth, Gurindar S. Sohi. Effective Jump-Pointer Prefetching for Linked Data Structures. ISCA, 1999 (26 citations)
Trishul M. Chilimbi, Mark D. Hill, James R. Larus. Cache-Conscious Structure Layout. PLDI, 1999 (54 citations)
Shai Rubin, David Bernstein, Michael Rodeh. Virtual Cache Line: A New Technique to Improve Cache Exploitation for Recursive Data Structures. CC, 1999 (3 citations)
Brad Calder, Chandra Krintz, Simmi John, Todd M. Austin. Cache-Conscious Data Placement. ASPLOS, 1998 (27 citations)
Bibtex Entry: [Edit]
@inproceedings{cahoon1999tolerating,
author = "Brendon Cahoon and Kathryn S. McKinley",
title = "Tolerating Latency by Prefetching Java Objects",
booktitle = "To appear: Workshop on Hardware Support for Objects and Microarchitectures for Java",
institution = "Department of Computer Science, University of Massachusetts",
year = "1999" }
Topics:
cache (26.9%), experimental results (20.9%), memory (9%), object (6%), high (4.5%), java (4.5%), algorithms (4.5%), accuracy (4.5%), techniques (4.5%)
Grants: (1)
James F. Kurose, John A. Stankovic, Donald F. Towsley, Krithi Ramamritham, J. Eliot B Moss, W. Richards Adrion, W. Bruce Croft, Kathryn McKinley. CISE Research Infrastructure: Infrastructure to Support Research on Networked Multimedia Information Systems. NSF EIA, 1995
Done
Adblock


Rexa: Kurose August 1, 1995 CISE Research Infrastructure: Infrastructure to Support Research on Networked Multimedi...
http://rexa.info/grant?id=8F07FF9BAB69A62C95048A2EFC6A2BA6F25403I
Rexa.info
Research • People × Connections
Andrew McCallum · Tags · Send Invites (477) · Submit · Logout
Papers Authors Grants
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
CISE Research Infrastructure: Infrastructure to Support Research on Networked Multimedia Information Systems [Google]
James F. Kurose, John A. Stankovic, Donald F. Towsley, Krithi Ramamritham, J. Eliot B Moss, W. Richards Adrion, W. Bruce Croft, Kathryn McKinley
NSF Grant EIA-9502639, August 1, 1995 - December 29, 1999
Abstract:
This award provides support to equip a networked, experimental testbed to enable research in the development of the operating system, I/O, networking, object management, and information retrieval components of future networked multimedia information systems. The testbed will consist of two shared-memory multiprocessor facilities attached to several parallel mass storage I/O devices and a high-speed ATM network. The research team will be developing several key hardware and software technologies needed to support future networked, multimedia information systems. Specific research areas include operating systems, I/O, networking, object management and information retrieval.
Papers: (17) Sorted by date | citations | alphabetically
This may be only a partial list of papers for this grant.
• Emery D. Berger, Benjamin G. Zorn, Kathryn S. McKinley. Composing High-Performance Memory Allocators. PLDI, 2001 (7 citations)
• Brendon Cahoon, Kathryn S. McKinley. Data Flow Analysis for Software Prefetching Linked Data Structures in Java. IEEE PACT, 2001 (11 citations)
• Sally Floyd, Mark Handley, Jitendra Padhye, Jörg Widmer. Equation-based congestion control for unicast applications. SIGCOMM, 2000 (229 citations)
• Sally Floyd, Mark Handley, Jitendra Padhye. Equation-Based Congestion Control for Unicast Applications \Lambda. 2000 (7 citations)
• Supratik Bhattacharyya, Don Towsley, James F. Kurose. Design and Analysis of Loss Indication Filters for Multicast Congestion Control. CMPSCI Technical Report TR 99-46, Department of Computer Science University of Massachusetts Amherst, 2000 (0 citations)
• Kathryn S. McKinley, Olivier Temam. Quantifying loop nest locality using SPEC'95 and the perfect benchmarks. ACM Trans. Comput. Syst. vol 17, pages 288, 1999 (9 citations)
• Brendon Cahoon, Kathryn S. McKinley. Tolerating Latency by Prefetching Java Objects. To appear: Workshop on Hardware Support for Objects and Microarchitectures for Java, 1999 (3 citations)
• Jitendra Padhye, James F. Kurose, Donald F. Towsley, Rajeev Koodli. A TCP-Friendly Rate Adjustment Protocol for Continuous Media Flows over Best Effort Networks CMPSCI


Rexa
http://rexa.info/results?pager.offset=0&pager.offset=0&searchIndex=autho
Rexa.info
Research • People × Connections
Andrew McCallum • Tags • Send Invites (477) • Submit • Logout
Papers Authors Grants
"machine learning" AND "reinforcement learning"
Search
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.
Search among authors using query "machine learning" AND "reinforcement learning"
Results 1-10 of about 307
1. Richard S. Sutton
editor. A Special Issue of Machine Learning on Reinforcement Learning, volume 8
Two problems with backpropagation and other steepestdescent learning procedures for networks
Open Theoretical Questions in Reinforcement Learning
editor
Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning
2. Thomas G. Dietterich
Divide and Conquer Methods for Machine Learning
Presidential Young Investigator Award (Computer and Information Science
Develop and Protype Methods for the Automatic Calibration and Validation of Computer Models of Complex Systems
Off-the-shelf Learning Algorithms for Structural Supervised Learning
Understanding and Scaling-Up Machine Learning Algorithms
3. Andrew G. Barto
Lyapunov Methods for Reinforcement Learning
Associative search network: a reinforcement learning associative memory
If desired, the present analysis of the stochastic neuronal dynamics can be replaced by an analysis of this deterministic neuronal dynamics
An approach to Learning Control Surfaces by Connectionist Systems
Reinforcement learning and its relationship to supervised learning
4. Martin A. Riedmiller
Learning to control dynamic systems
Aspects of learning neural control
High quality thermostat control by reinforcement learning - a case study
Karlsruhe Brainstormers - Design Principles

Rexa: "information retrieval" topic

http://rexa.info/htdocs/topics/topic25.html

Rexa.info
Research • People × Connections

Papers Authors Grants
Search Advanced Help
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.

Topic: "information retrieval"

Citations to this topic: **19697** (rank **59/400**)
Impact diversity: **2.91** (rank **387/400**)

Topic terms:

Words		Phrases	
0.1290	retrieval	0.1844	information retrieval
0.0600	documents	0.0773	relevance feedback
0.0569	document	0.0761	image retrieval
0.0469	indexing	0.0398	query expansion
0.0469	information	0.0380	text retrieval
0.0463	content	0.0336	search engines
0.0391	query	0.0282	search engine
0.0273	relevance	0.0240	image databases
0.0242	collection	0.0208	latent semantic indexing
0.0241	search	0.0197	relevant documents

Citing topics

- experimental results (3877)
- text (633)
- web (610)
- query language (481)
- word (415)
- video (296)
- image (257)
- search (242)
- semantic web (217)
- information (217)
- user (199)

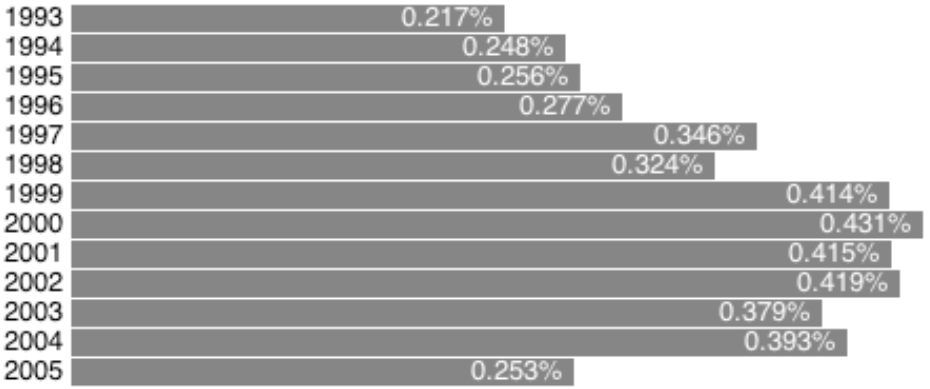
Cited topics

- information (1231)
- experimental results (1085)
- text (705)
- web (636)
- search (558)
- word (547)
- query language (434)
- image (415)
- access (289)
- world wide web (287)
- neural networks (214)

Cooccurring topics

- word (0.03411)
- experimental results (0.03019)
- image (0.02958)
- text (0.02817)
- web (0.02627)

Trends: papers | **% of all papers** | citations | % of all cites (recent coverage sparse)



Top papers: Sorted by **citations** | broadest impact | earliest

- Gerard Salton, Chris Buckley. ***Term-Weighting Approaches in Automatic Text Retrieval***. (257 citations)
- Myron Flickner, Harpreet S Sawhney, Jonathan J Ashley, Qiang Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, Peter Yanker. ***Query by Image and Video Content: The QBIC System***. (250 citations)
- Douglas R Cutting, Jan O Pedersen, David R Karger, John W Tukey. ***Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections***. (140 citations)
- Wayne Niblack, Ron Barber, William Equitz, Myron Flickner, Eduardo H Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, Gabriel Taubin. ***The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape***. (137 citations)
- A Pentland, R Picard, S Sclaroff. ***Photobook: Content-based***

Rexa: "neural networks" topic

http://rexa.info/htdocs/topics/topic392.html

Rexa.info
Research • People × Connections

Papers Authors Grants
Search Advanced Help
Optional fields include abstract: body: title: author: venue: year: tag:
Queries may use AND, OR or (). Default is OR.

Topic: "neural networks"

Citations to this topic: **28048** (rank **22/400**)
Impact diversity: **3.66** (rank **278/400**)

Topic terms:

Words		Phrases	
0.0955	neural	0.3318	neural networks
0.0908	learning	0.1565	neural network
0.0837	training	0.0425	artificial neural networks
0.0404	network	0.0227	organizing maps
0.0365	recurrent	0.0214	associative memory
0.0360	networks	0.0171	neural nets
0.0313	organizing	0.0168	organizing map
0.0253	trained	0.0163	hidden units
0.0222	connectionist	0.0125	artificial neural network
0.0198	weights	0.0112	recurrent networks

Citing topics

- experimental results (9332)
- classification (805)
- learning (709)
- visual cortex (614)
- basal ganglia (557)
- cognitive (397)
- bayesian (384)
- university (351)
- mobile robot (334)
- genetic algorithms (321)
- speech recognition (290)

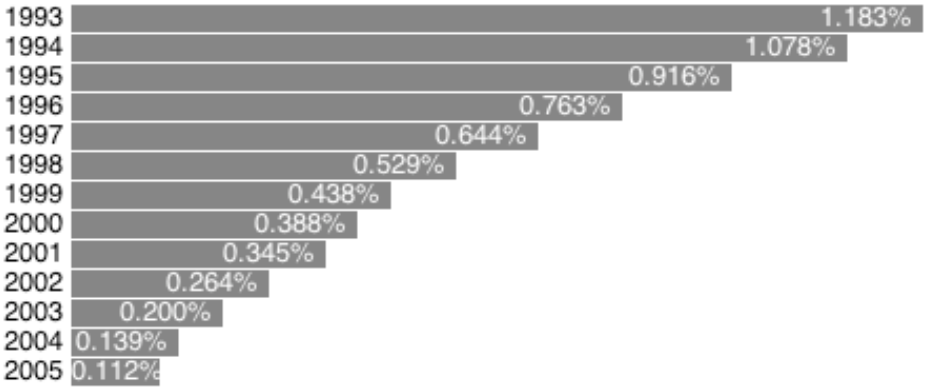
Cited topics

- experimental results (656)
- visual cortex (499)
- cognitive (411)
- basal ganglia (410)
- learning (387)
- error (287)
- speech recognition (252)
- curves (228)
- breast cancer (218)
- bayesian (183)
- recognition (183)

Cooccurring topics

- fuzzy (0.01314)
- genetic algorithms (0.01227)
- de (0.01125)
- recognition (0.01102)
- features (0.01024)

Trends: papers | **% of all papers** | citations | % of all cites (recent coverage sparse)



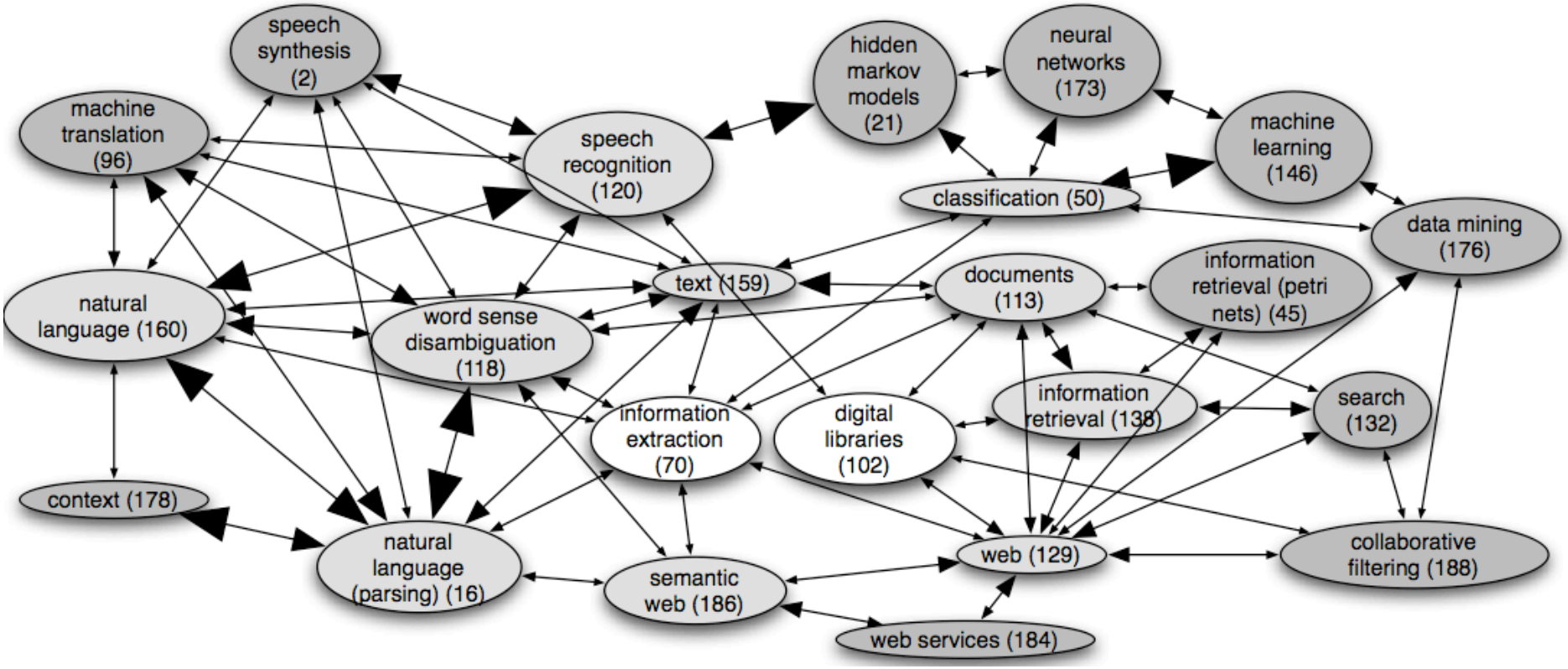
Top papers: Sorted by **citations** | broadest impact | earliest

- Kurt Hornik, Maxwell Stinchcombe, Halbert White. ***Multilayer Feed-forward Neural Networks Are Universal Approximators,***. (235 citations)
- Howard A Rowley, Shumeet Baluja, Takeo Kanade. ***Neural Network-Based Face Detection***. (197 citations)
- Stuart Geman, Elie Bienenstock, R Doursat. ***Neural networks and the bias/variance dilema***. (167 citations)
- Teuvo Kohonen. ***The self-organizing map***. (163 citations)
- Scott E Fahlman, Christian Lebiere. ***The Cascade-Correlation Learning Architecture***. (147 citations)
- Anders Krogh, Jesper Vedelsby. ***Neural Network Ensembles, Cross Validation, and Active Learning***. (101 citations)
- P Tamayo. ***Interpreting patterns of gene expression with self-organizing maps: methods and application,***. (100 citations)

Topical Transfer

Citation counts from one topic to another.

Map "producers and consumers"



Topical Bibliometric Impact Measures

[Mann, Mimno, McCallum, 2006]

- Topical Citation Counts
- Topical Impact Factors
- Topical Longevity
- Topical Precedence
- Topical Diversity
- Topical Transfer

Topical Transfer

Transfer from Digital Libraries to other topics

Other topic	Cit's	Paper Title
Web Pages	31	*Trawling the Web for Emerging Cyber-Communities,* Kumar, Raghavan,... 1999.
Computer Vision	14	*On being 'Undigital' with digital cameras: extending the dynamic...*
Video	12	*Lessons learned from the creation and deployment of a terabyte digital video libr..*
Graphs	12	*Trawling the Web for Emerging Cyber-Communities*
Web Pages	11	*WebBase: a repository of Web pages*

Topical Diversity

Papers that had the most influence across many other fields...

Topical Diversity	Citations	Title
4.00	618	A tutorial on hidden Markov models and selected applications in speech processing
3.80	138	The self-organizing map
3.77	163	Hierarchical mixtures of experts and the EM algorithm
3.74	65	Quantifying Inductive Bias: AI Learning Algorithms and ...
3.74	144	Knowledge Acquisition via Incremental Conceptual Clustering
3.73	155	A Tutorial on Learning With Bayesian Networks
3.72	244	Term-Weighting Approaches in Automatic Text Retrieval
3.71	294	Finding Structure in Time
3.7	173	An introduction to hidden Markov models
3.7	132	Nearest neighbor pattern classification

Topical Diversity

Entropy of the topic distribution among papers that cite this paper (this topic).

Topic	Impact Diversity
Simulated Annealing (52)	4.59
Pattern Recognition (125)	4.57
Probabilistic Modeling (3)	4.55
Finite Automata (66)	4.55
Probability (89)	4.5
Digital Libraries (102)	3.77
Machine Translation (96)	3.32
Mobile Robots (22)	3.31
Graphics (9)	3.21
Speech Recognition (120)	3.09
Computer Vision (49)	2.95

High Diversity

Low Diversity

Topical Bibliometric Impact Measures

[Mann, Mimno, McCallum, 2006]

- Topical Citation Counts
- Topical Impact Factors
- Topical Longevity
- Topical Precedence
- Topical Diversity
- Topical Transfer

Topical Precedence

"Early-ness"

Within a topic, what are the earliest papers that received more than *n* citations?

Speech Recognition:

Some experiments on the recognition of speech, with one and two ears,
E. Colin Cherry (1953)

Spectrographic study of vowel reduction,
B. Lindblom (1963)

Automatic Lipreading to enhance speech recognition,
Eric D. Petajan (1965)

Effectiveness of linear prediction characteristics of the speech wave for...,
B. Atal (1974)

Automatic Recognition of Speakers from Their Voices,
B. Atal (1976)

Topical Precedence

"Early-ness"

Within a topic, what are the earliest papers that received more than *n* citations?

Information Retrieval:

On Relevance, Probabilistic Indexing and Information Retrieval,
Kuhns and Maron (1960)

Expected Search Length: A Single Measure of Retrieval Effectiveness Based on the Weak Ordering Action of Retrieval Systems,
Cooper (1968)

Relevance feedback in information retrieval,
Rocchio (1971)

Relevance feedback and the optimization of retrieval effectiveness,
Salton (1971)

New experiments in relevance feedback,
Ide (1971)

Automatic Indexing of a Sound Database Using Self-organizing Neural Nets,
Feiten and Gunzel (1982)

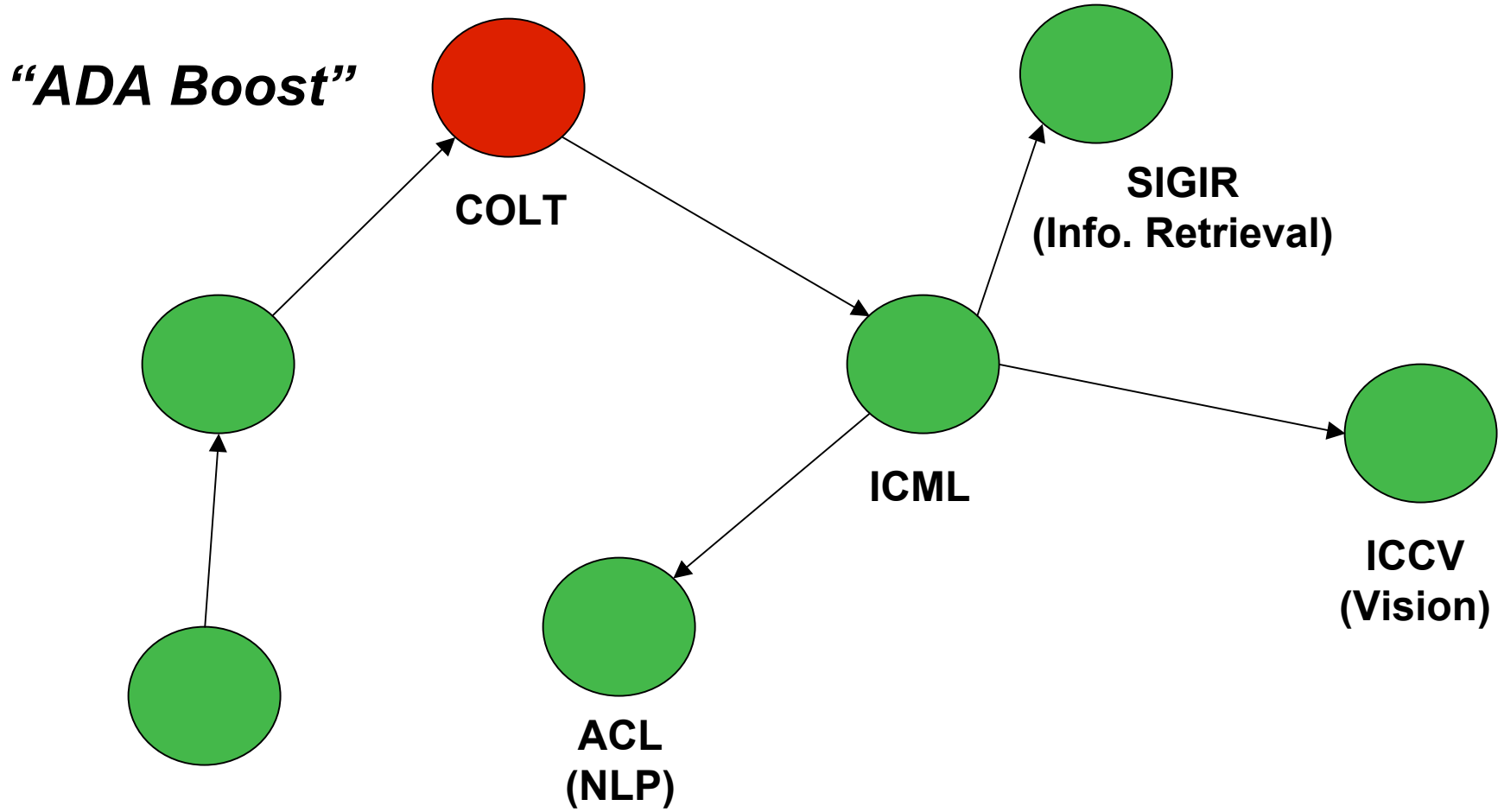
Topical Transfer Through Time

- Can we predict which research topics will be “hot” at ICML *next year*?

- ...based on
  - the hot topics in “neighboring” venues *last year*
  - learned “neighborhood” distances for venue pairs

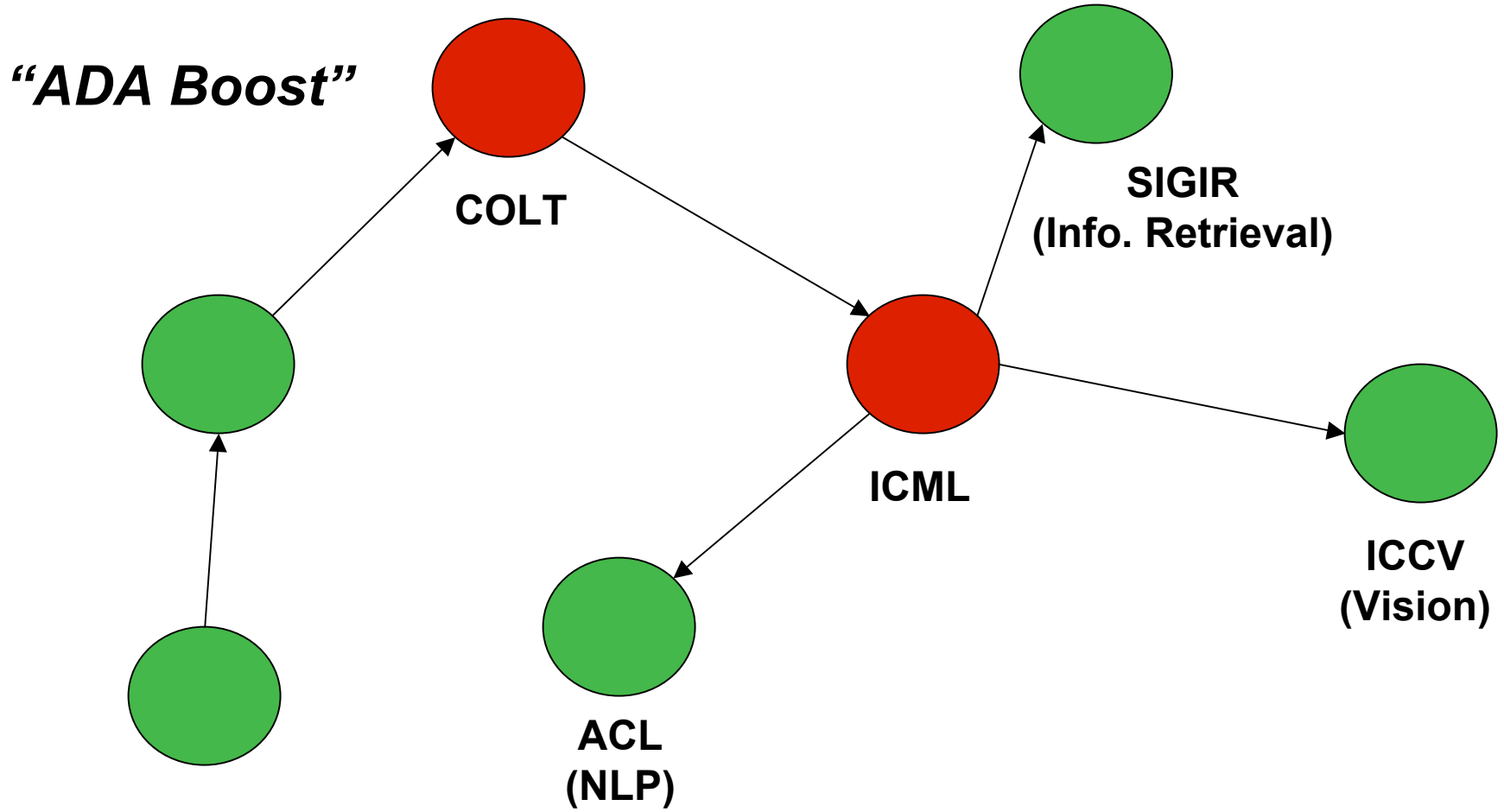
How do Ideas Progress Through Social Networks?

Hypothetical Example:



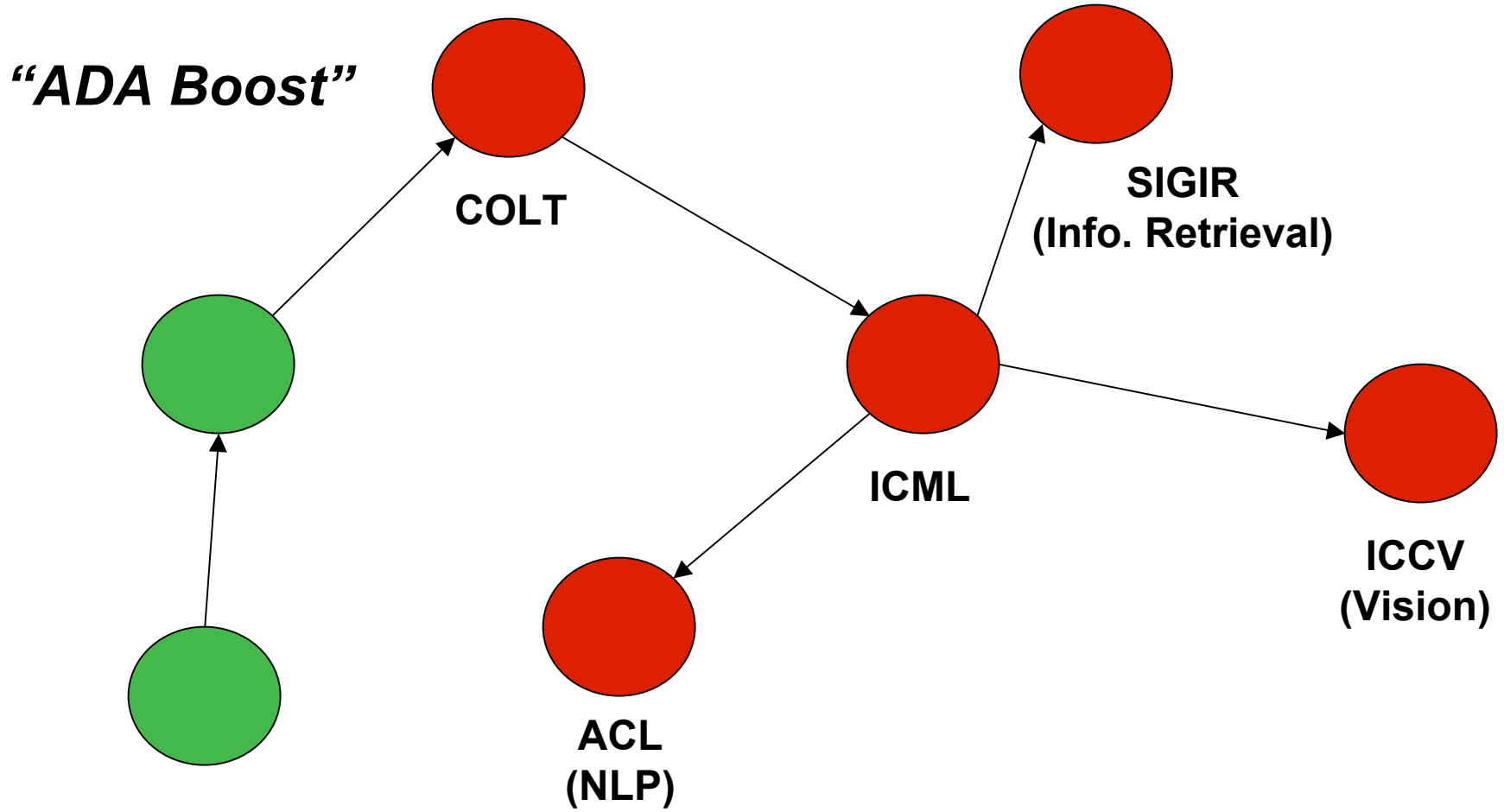
How do Ideas Progress Through Social Networks?

Hypothetical Example:



How do Ideas Progress Through Social Networks?

Hypothetical Example:



Topic Prediction Models

Static Model $$Z_v^i = \lambda_v$$

Transfer Model $$Z_v^i = \lambda_v + \sum_{v'} \theta_v^{v'} Z_{v'}^{i-1}$$

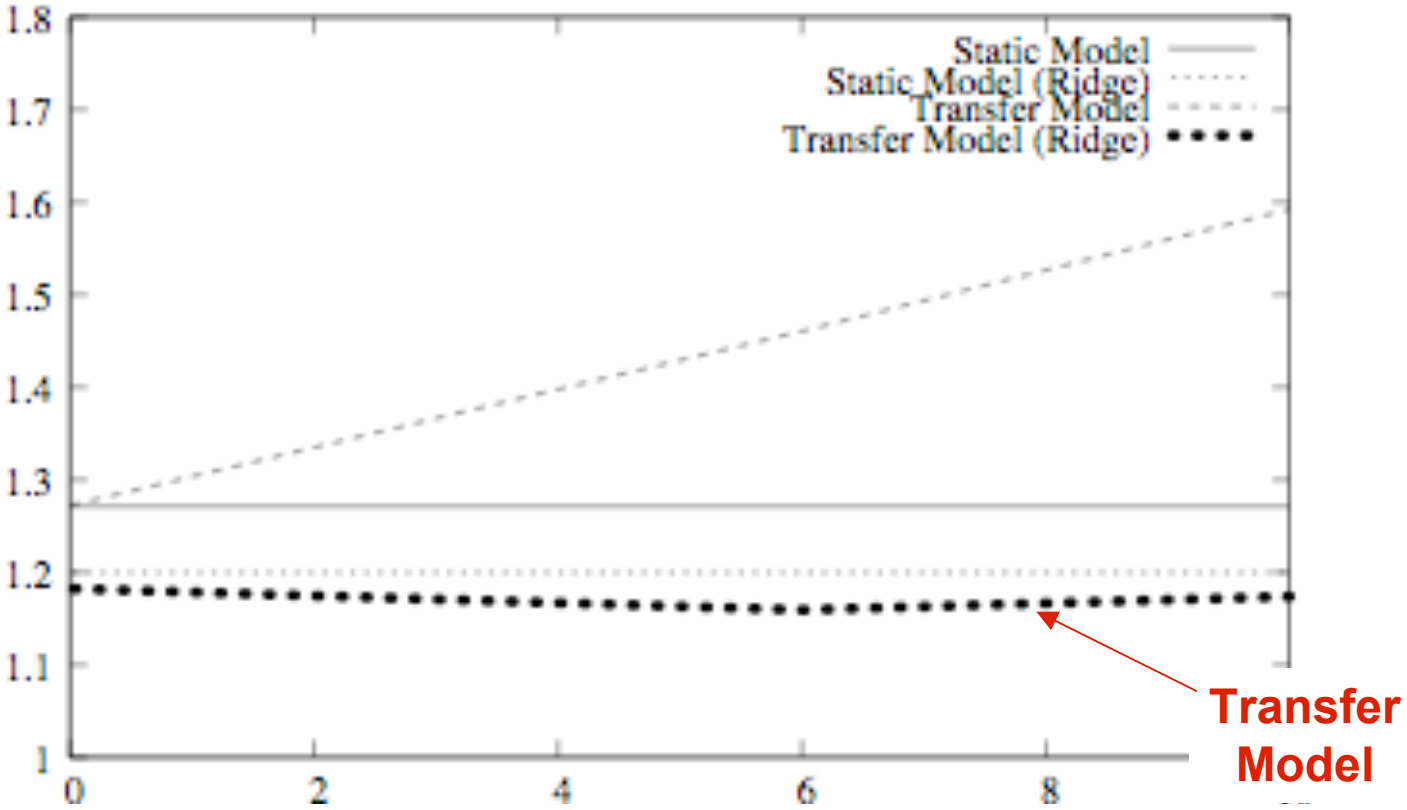
- Z_v^i : proportion of topic Z in venue v in year i
- λ_v : static topic coefficient
- $\theta_v^{v'}$: topic transfer coefficient

Linear Regression and Ridge Regression
Used for Coefficient Training.

Preliminary Results

Mean Squared Prediction Error

(Smaller Is better)



Venues used for prediction

Transfer Model with Ridge Regression is a good Predictor

Outline

- Social Network Analysis
  - Roles *(Author-Recipient-Topic Model)*
  - Groups *(Group-Topic Model)*
  - Trends over time *(Topics-over-Time Model, TOT)*
  - Preferential Attachment *(Community-Author-Topic, CAT)*
- Undirected Graphical Models
  - Flexible Objective Functions *(Multi-Conditional Learning, MCL)*
  - Topics for Prediction *(Multinomial-Components-Analysis, MCA)*
- Demo: *Rexa*, a Web portal for researchers
  - Topical Impact Measures (Diversity,...)

Social Network Analysis
(Pattern Discovery in Networks)

Data:

Network Connectivity → **Network Connectivity + *Many Attributes***

Text, Timestamps, Authors,...

Objective:

Descriptive → **Predictive & Prescriptive**

Small World, Betweeness Centrality,...

P(edge|...) P(attribute|...), P(group|...)
P(collaborator|...)

Methodology:

Direct Measures, Agglomerative & Spectral Clustering,... → **Generative, Latent Variable Models**