Reduced Rank Factor Modeling

Ayan Acharya

October 8, 2015

Matrices and Tensors are Ubiquitous

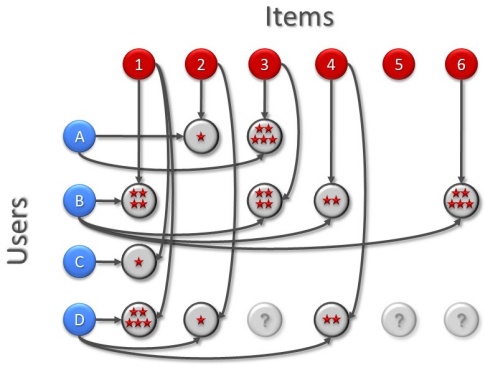


Figure: Movie Recommendation

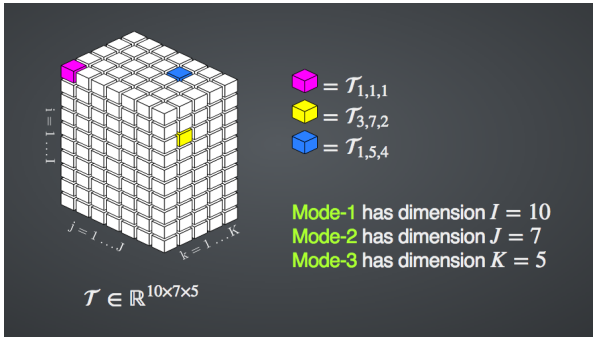
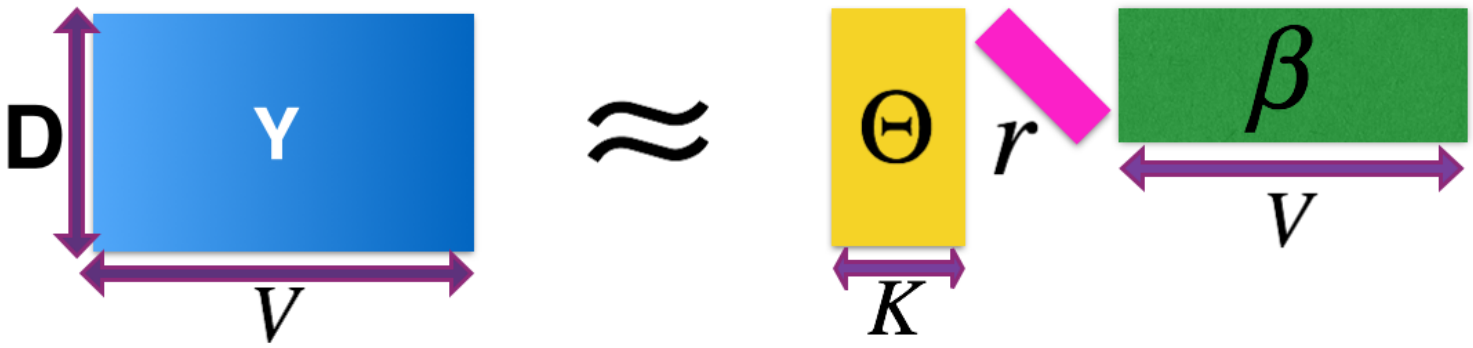


Figure: EHR Data

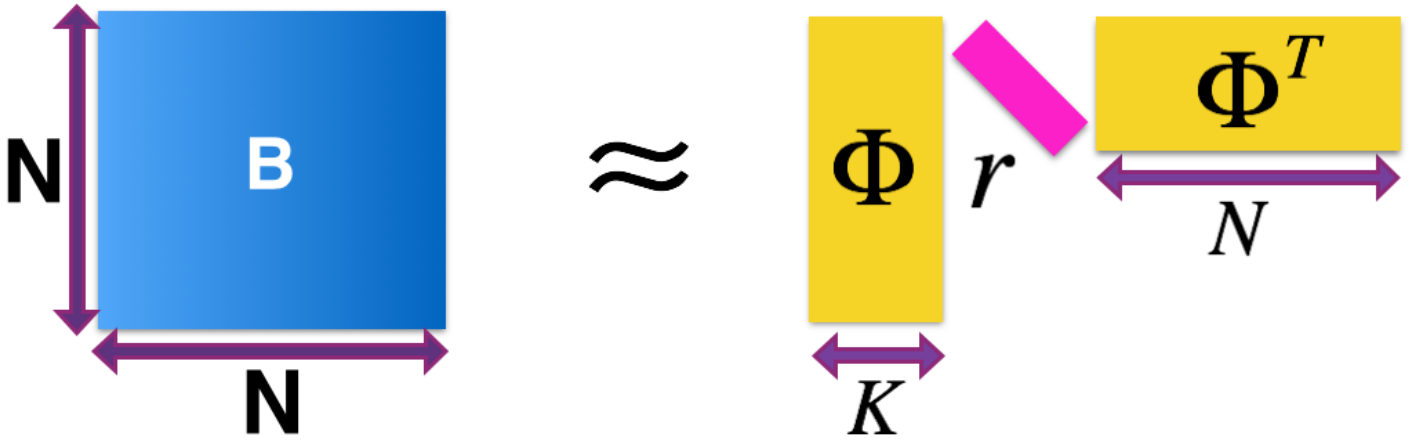
Figure: Social Network

Singular Value Decomposition



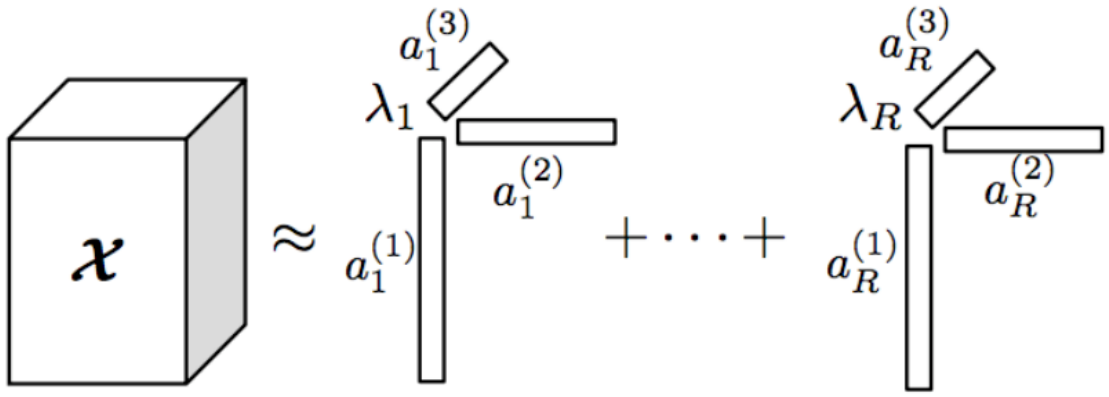
- $y_{dw} \approx \sum_{k=1}^{K} r_k \theta_{dk} \beta_{wk}$.

Eigen Value Decomposition



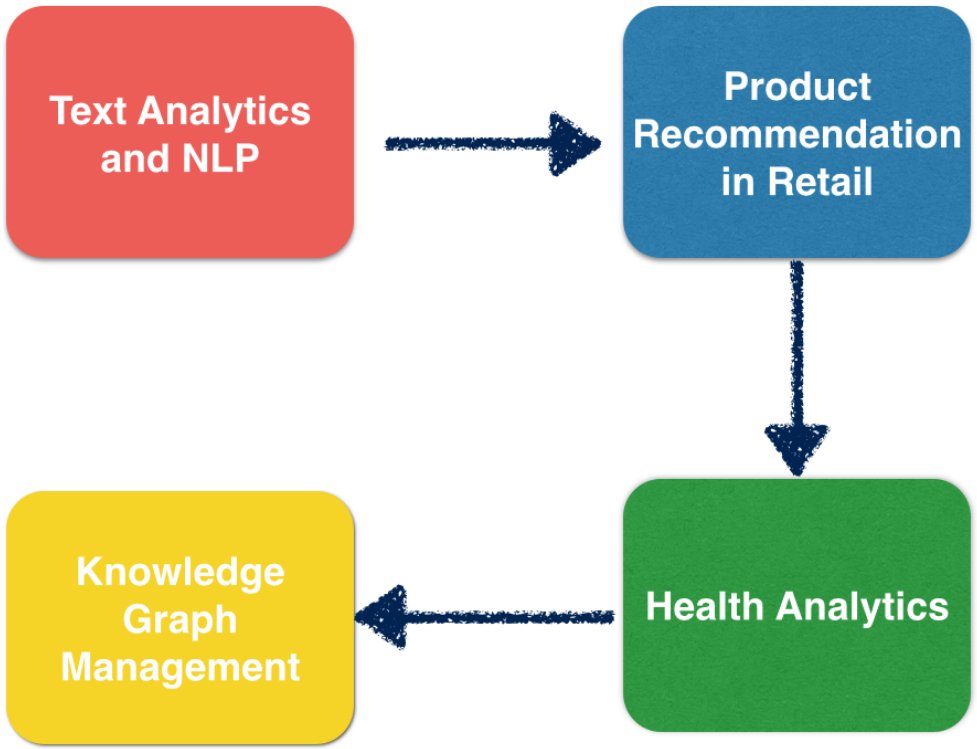
- $b_{nm} \approx \sum_{k=1}^{K} r_k \phi_{nk} \phi_{mk}$.

Tensor Decomposition



- $x_{d_1 d_2 d_3} \approx \sum_{k=1}^{K} \lambda_k a^{(1)}_{d_1 k} a^{(2)}_{d_2 k} a^{(3)}_{d_3 k}$.

Talk Outline



Latent Semantic Analysis

- Probabilistic Latent Semantic Analysis
- Latent Dirichlet Allocation (topic models)

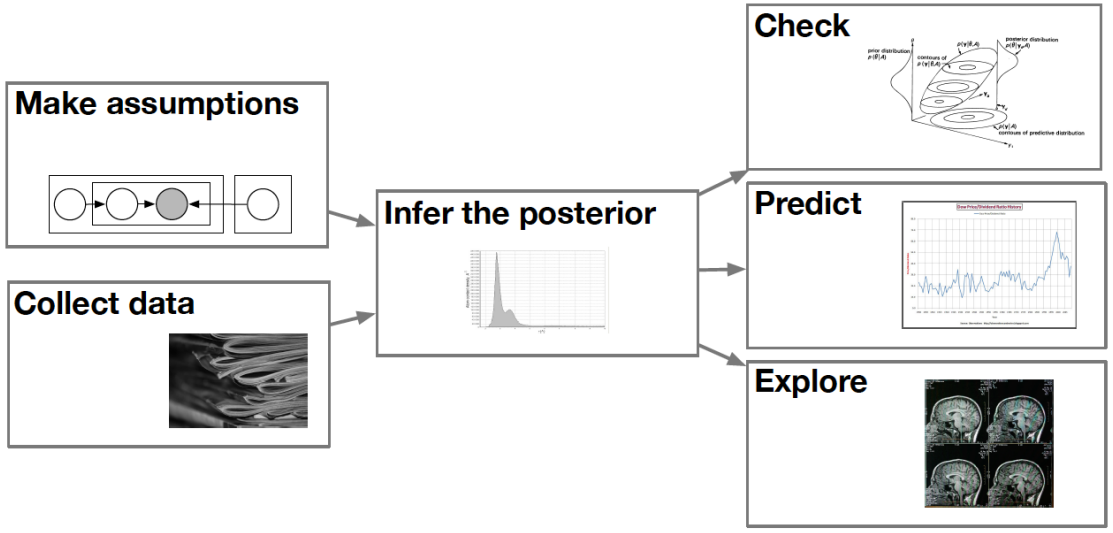


Figure: Objectives of Topic Models

Discovery of Topics (Insights?)

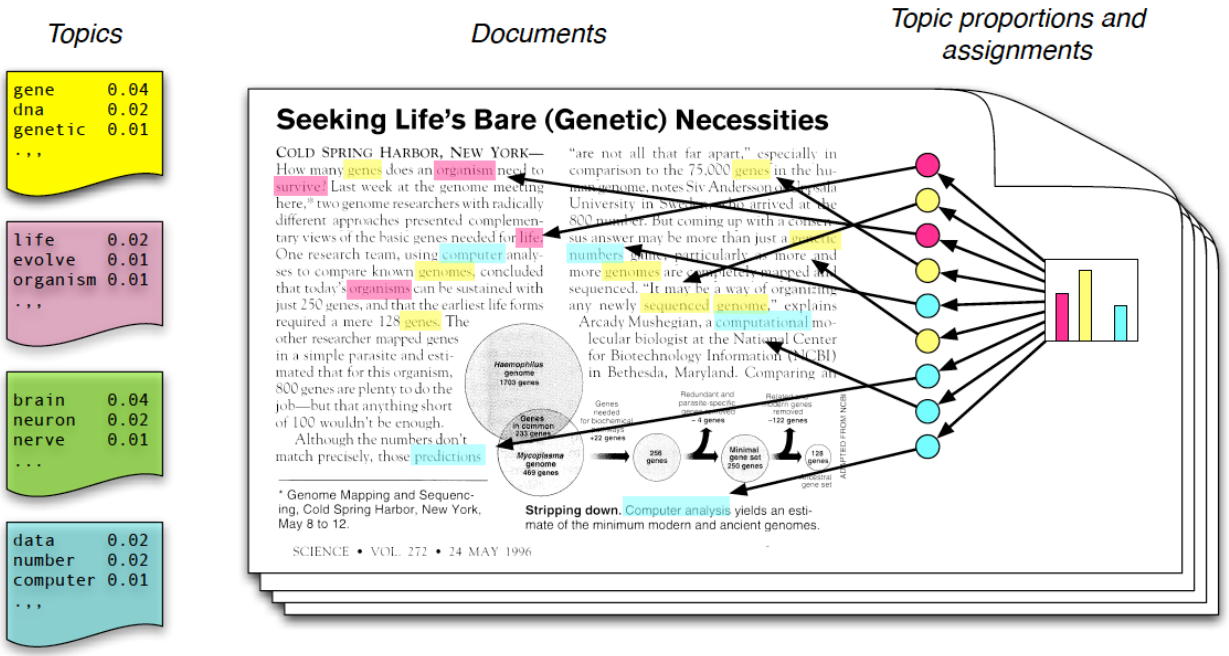


Figure: Assignment of an article to topics

Visualization of Topics

- Data: Collection of Science from 1990–2000, 17K documents, 11M words, 20K unique terms

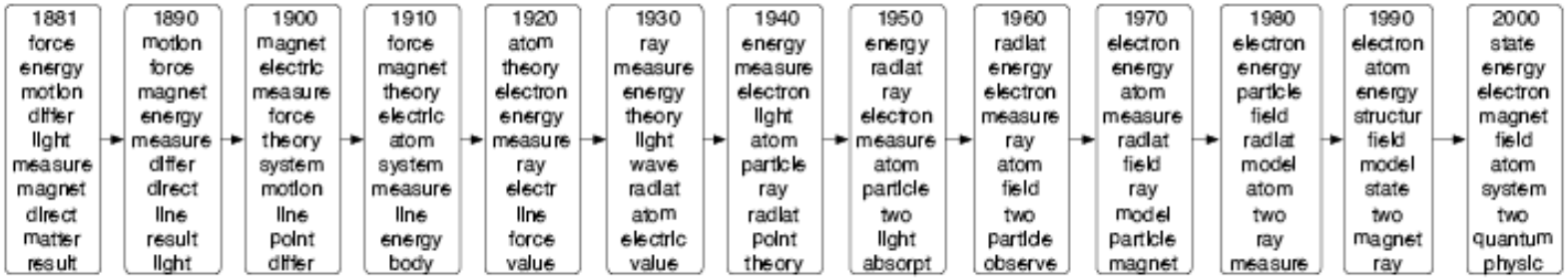
Domain Feedback into Topic Modeling

Operation	Meaning
Must-Link (school,college)	$\forall$ topics t, $P(school\|t) \approx P(college\|t)$
Cannot-Link (school,cure)	no topic t has $P(school\|t)$ and $P(cure\|t)$ both high

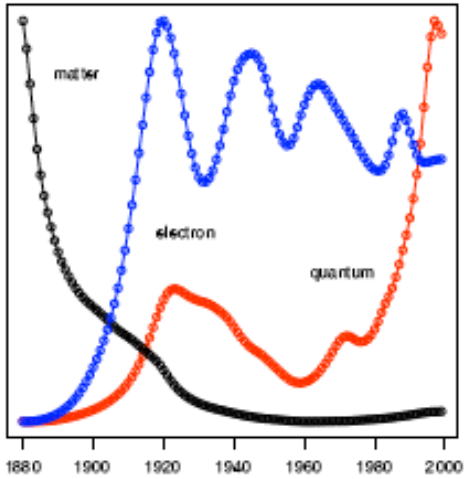
split	[*go school into college*] vs [*cancer free cure well*] $\rightarrow$ Must-Link among words for each concept $\rightarrow$ Cannot-Link between words from different concepts
merge	[*love marry together boyfriend*] in one topic [*married boyfriend engaged wedding*] in another $\rightarrow$ Must-Link among concept words
isolate	[*the year in 2008*] in many wish topics $\rightarrow$ Must-Link among words to be isolated $\rightarrow$ Cannot-Link vs other Top N words for each topic

Figure: Domain Feedback

What if the documents change with time?



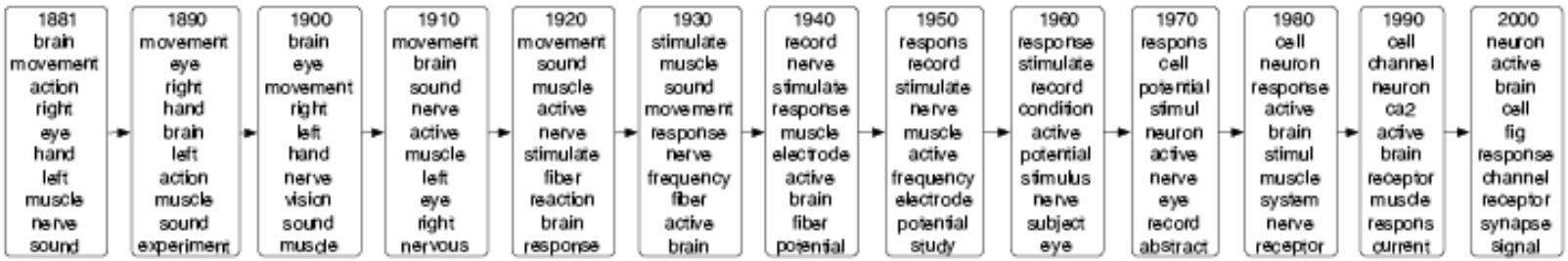
"Atomic Physics"



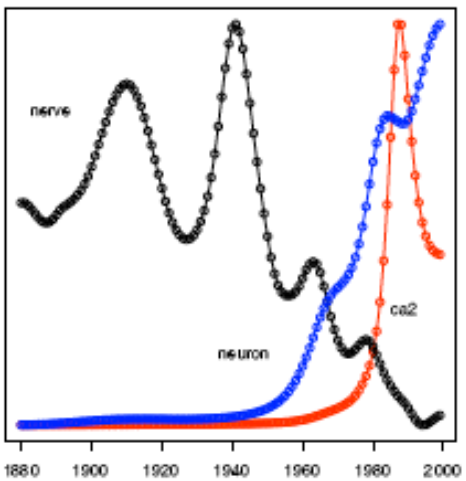
1881 On Matter as a form of Energy
1892 Non-Euclidean Geometry
1900 On Kathode Rays and Some Related Phenomena
1917 ``Keep Your Eye on the Ball"
1920 The Arrangement of Atoms in Some Common Metals
1933 Studies in Nuclear Physics
1943 Aristotle, Newton, Einstein. II
1950 Instrumentation for Radioactivity
1965 Lasers
1975 Particle Physics: Evidence for Magnetic Monopole Obtained
1985 Fermilab Tests its Antiproton Factory
1999 Quantum Computing with Electrons Floating on Liquid Helium

Figure: Evolution of Topic 1

What if the documents change with time?



"Neuroscience"



1887 Mental Science
1900 Hemianopsia in Migraine
1912 A Defence of the ``New Phrenology''
1921 The Synchronal Flashing of Fireflies
1932 Myoesthesis and Imageless Thought
1943 Acetylcholine and the Physiology of the Nervous System
1952 Brain Waves and Unit Discharge in Cerebral Cortex
1963 Errorless Discrimination Learning in the Pigeon
1974 Temporal Summation of Light by a Vertebrate Visual Receptor
1983 Hysteresis in the Force-Calcium Relation in Muscle
1993 GABA-Activated Chloride Channels in Secretory Nerve Endings

Figure: Evolution of Topic 2

Neural Word Embedding: Word2Vec (Deep Learning!)

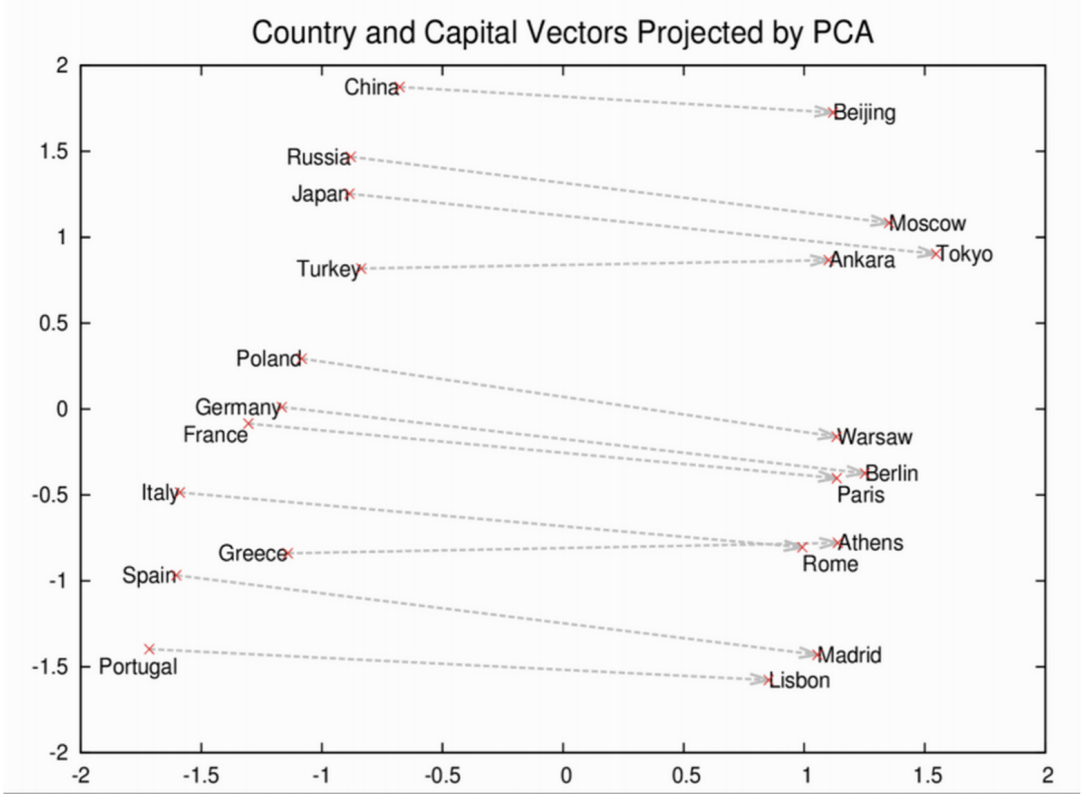


Figure: Illustration of Word2Vec

- http://www.ghostweather.com/files/word2vecpride/

Illustration of Word Embedding ..

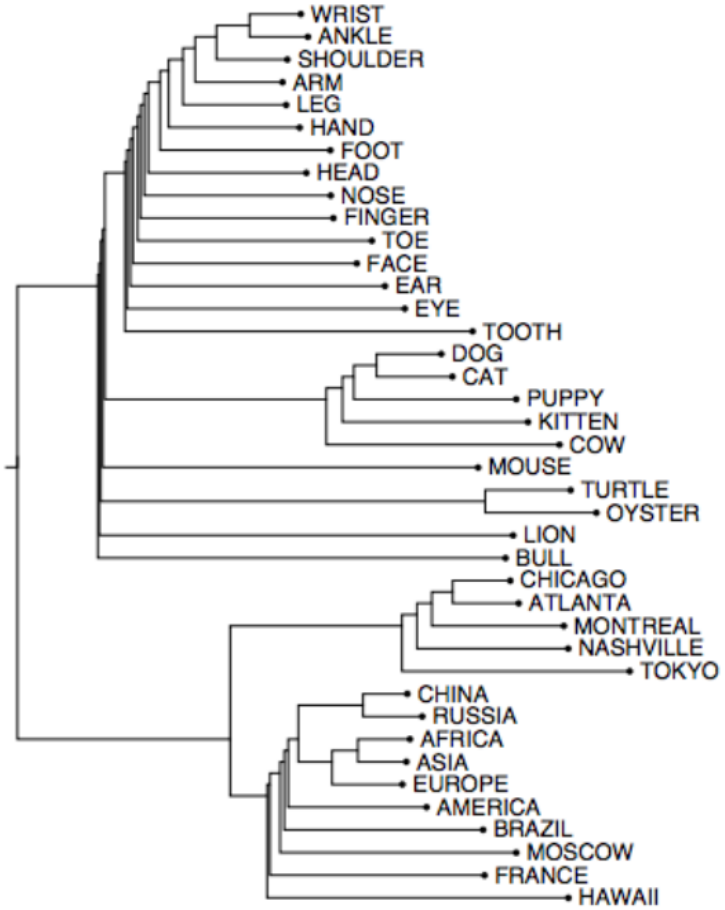


Figure: Illustration of Word Embedding

Illustration of Word Embedding

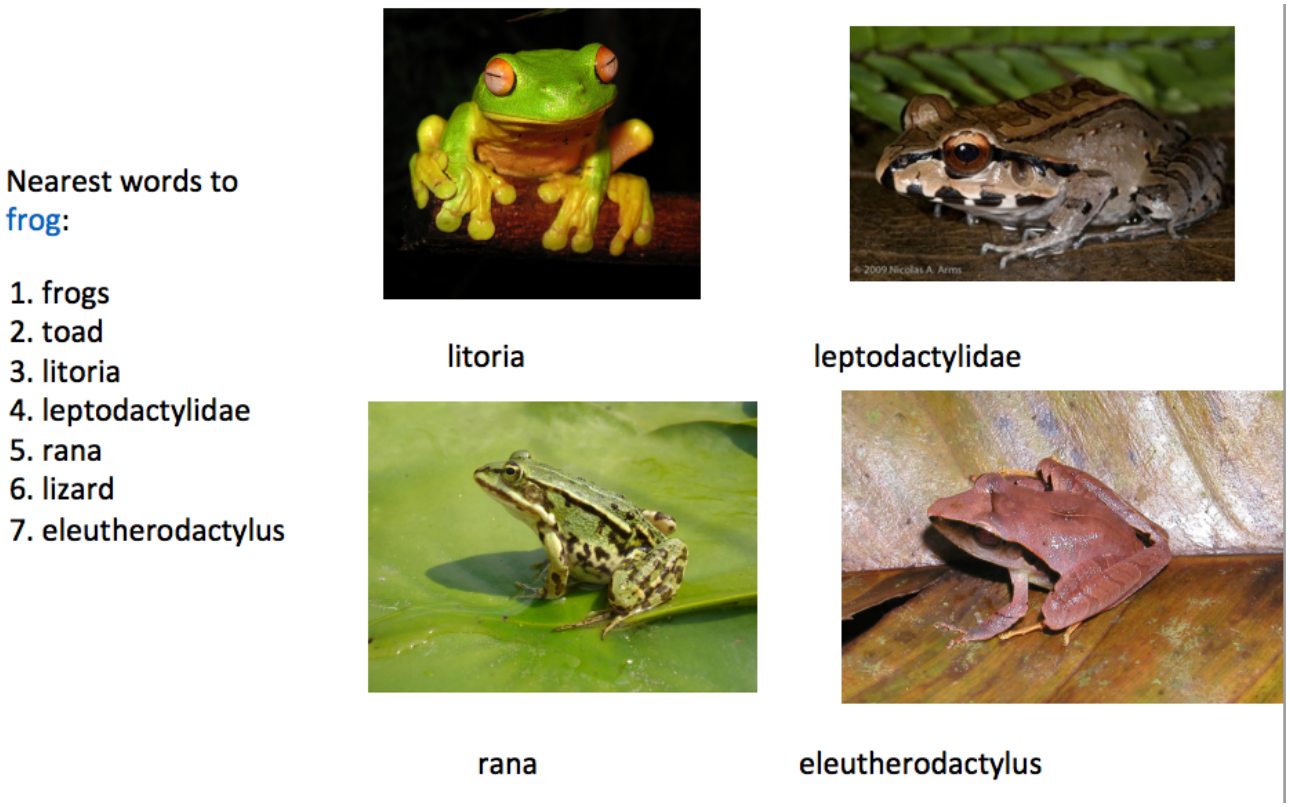


Figure: Illustration of Word Embedding

Recommender Systems

- Collaborative filtering
- Content-based filtering
- Content-based collaborative filtering

Musical instruments (Amazon)

drums	strings	wind	microphones	software
cartridge	guitar	reeds	mic	software
sticks	violin	harmonica	microphone	interface
strings	strap	cream	stand	midi
snare	neck	reed	mics	windows
stylus	capo	harp	wireless	drivers
cymbals	tune	fog	microphones	inputs
mute	guitars	mouthpiece	condenser	usb
heads	picks	bruce	battery	computer
these	bridge	harmonicas	filter	mp3
daddario	tuner	harps	stands	program

Video games (Amazon)

fantasy	nintendo	windows	ea/sports	accessories
fantasy	mario	sims	drm	cable
rpg	ds	flight	ea	controller
battle	nintendo	windows	spore	cables
tomb	psp	xp	creature	ps3
raider	wii	install	nba	batteries
final	gamecube	expansion	football	sonic
battles	memory	program	nhl	headset
starcraft	wrestling	software	basketball	wireless
characters	metroid	mac	madden	controllers
ff	smackdown	sim	hockey	component

Figure: Users' purchase pattern described by latent groups

What if the user-item interaction changes with time?

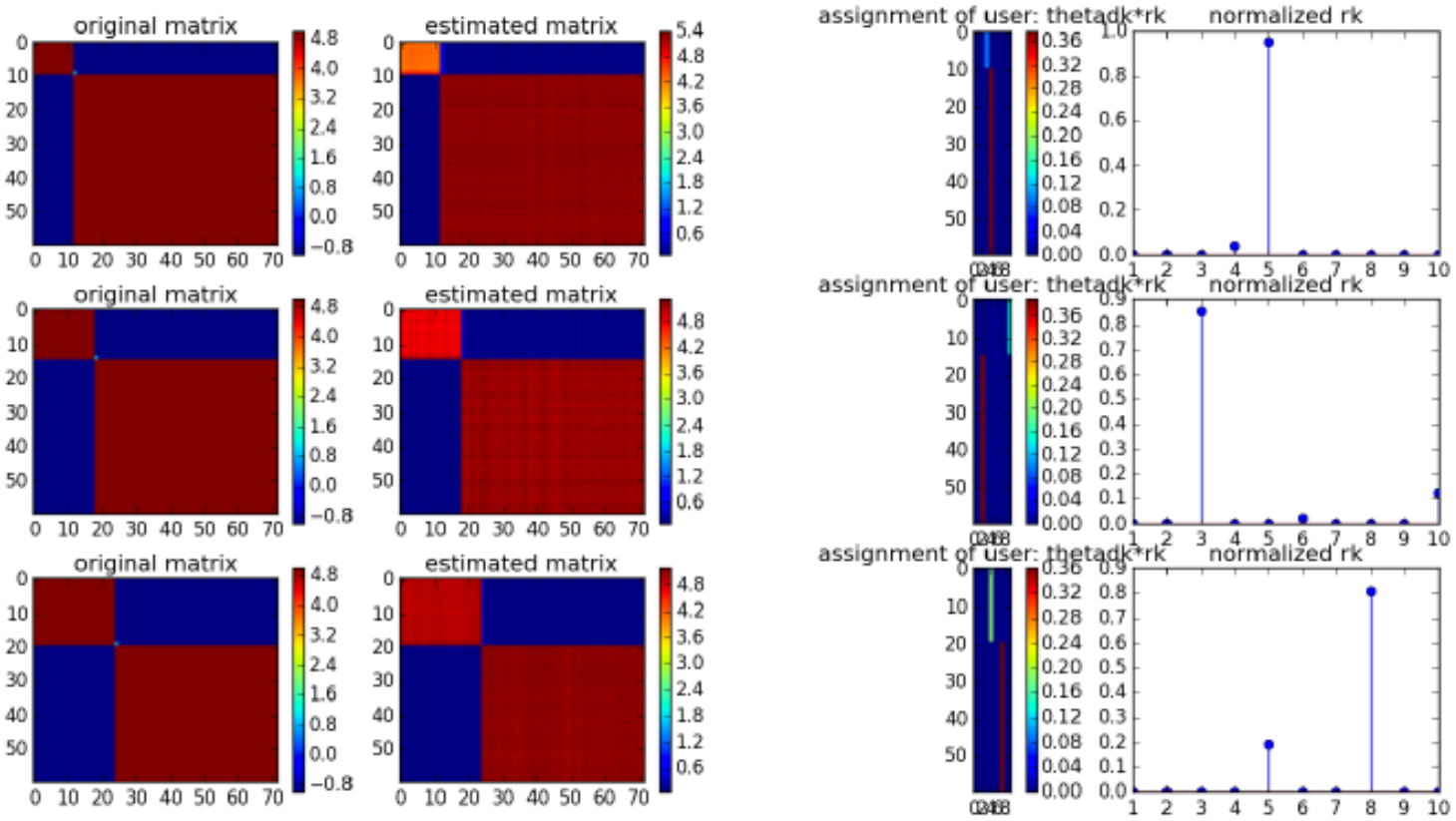


Figure: Static Recommendation

What if the user-item interaction changes with time?

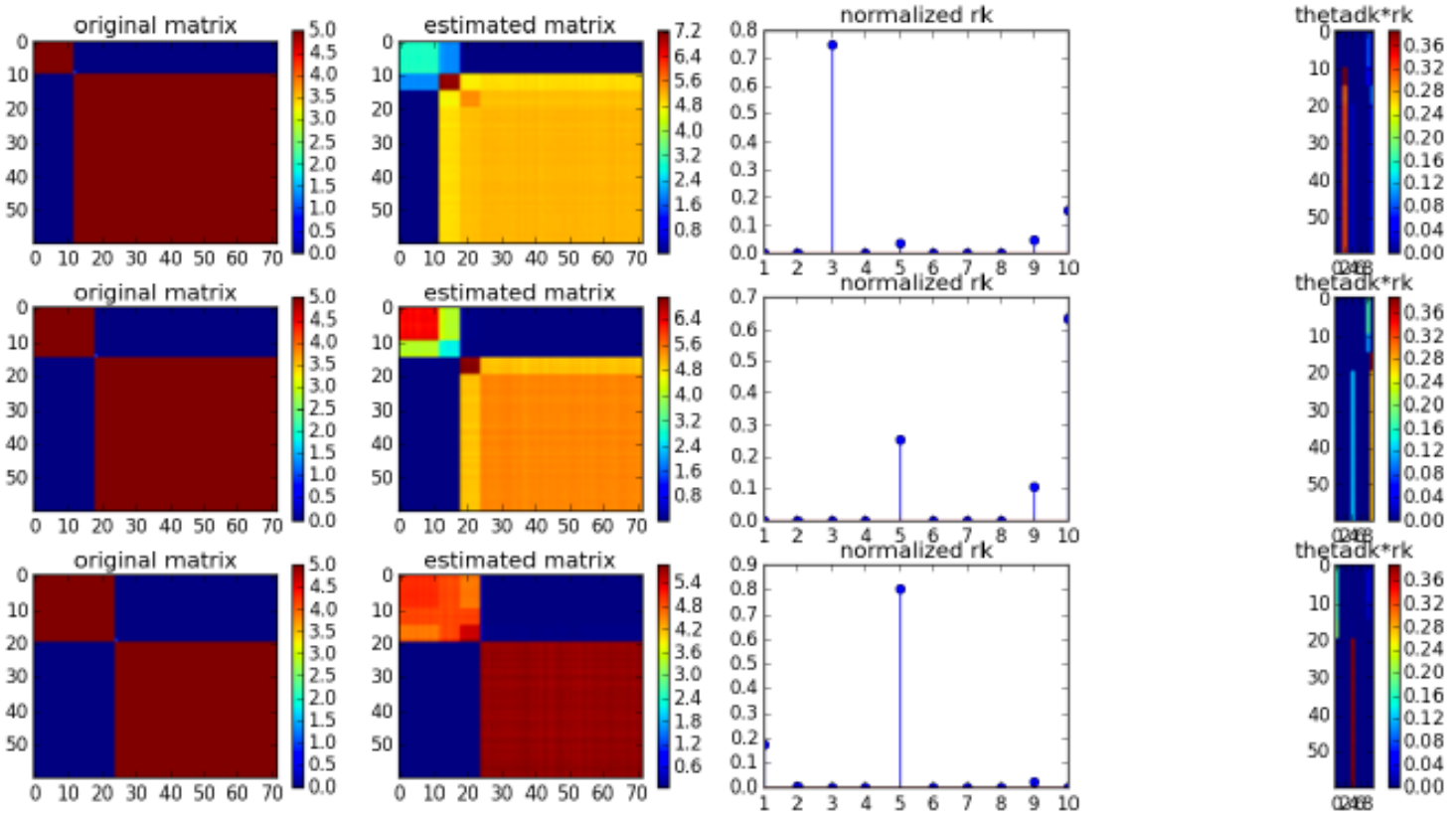


Figure: Dynamic Recommendation

Motivation – Manual Phenotyping of EHR Data

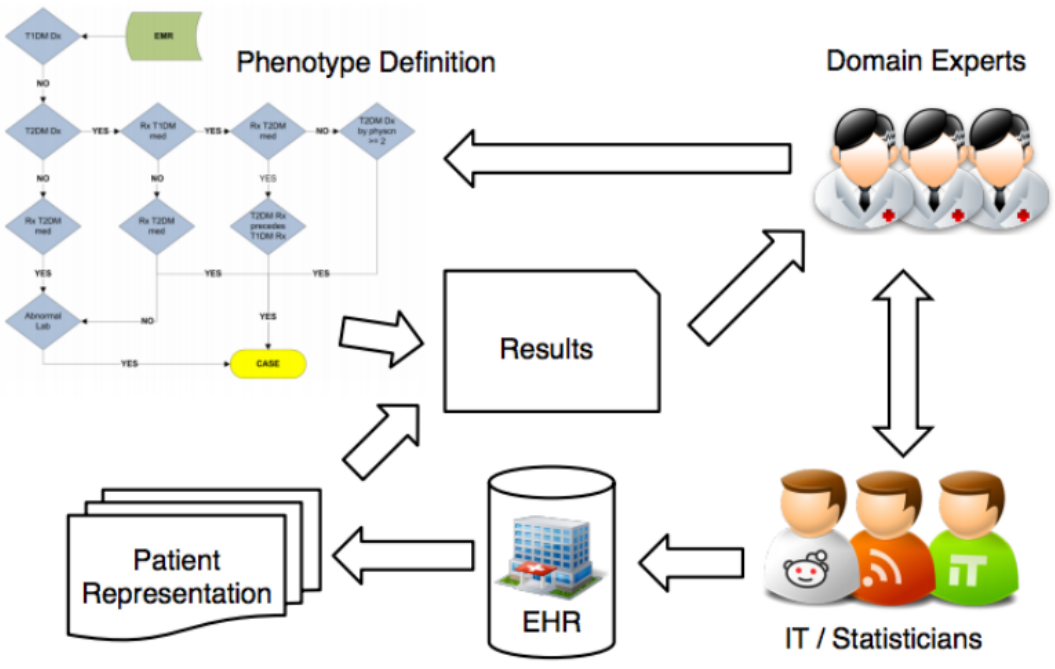


Figure: Manual Phenotyping of EHR Data

Automated Phenotyping

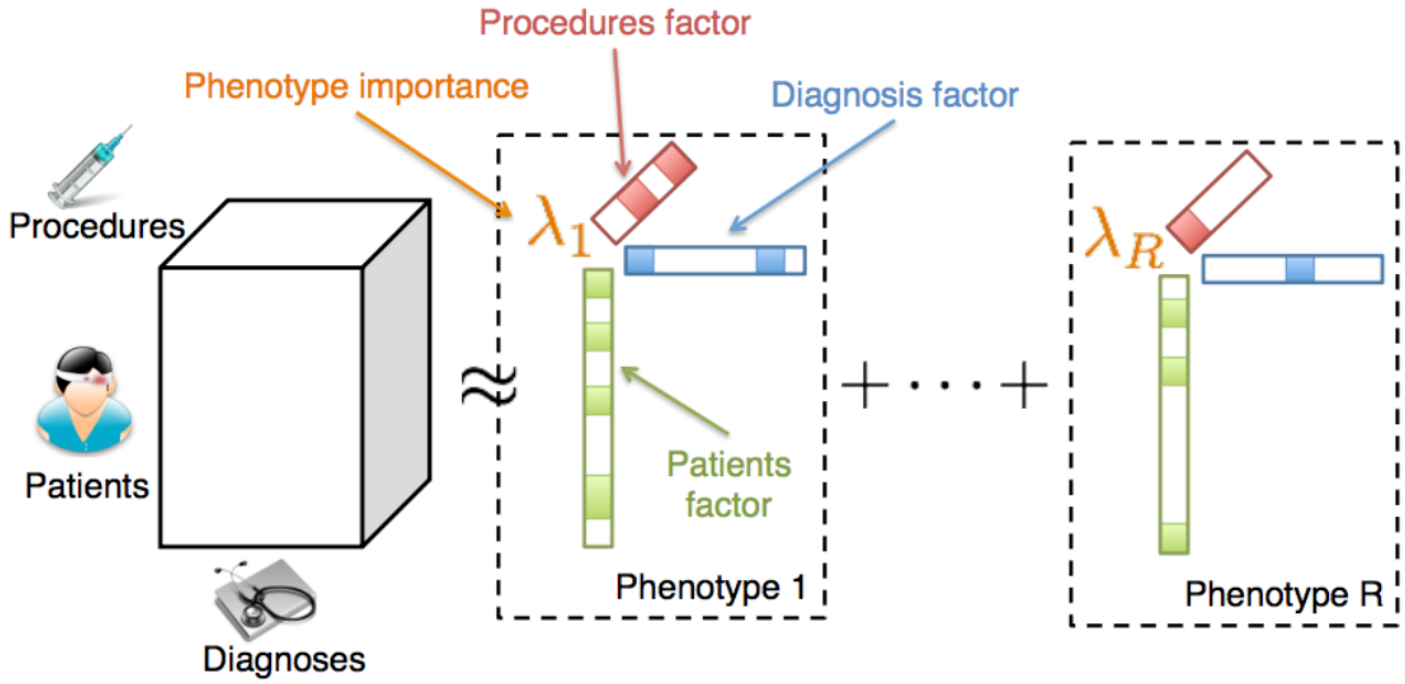


Figure: Automated Phenotyping of EHR Data

Phenotypes Discovered

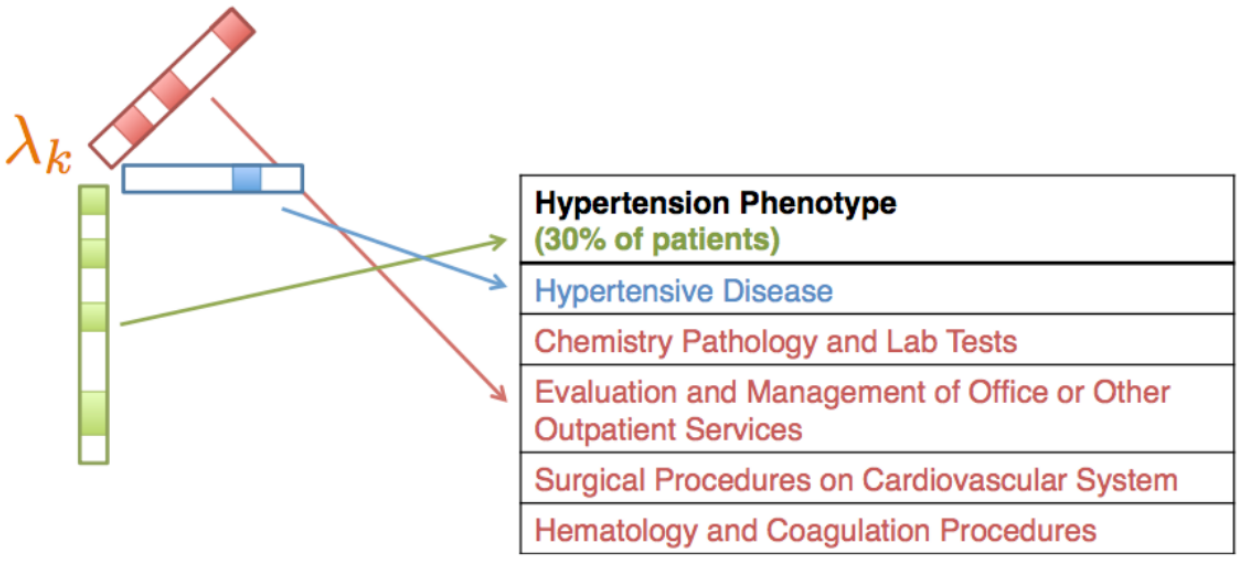


Figure: Phenotypes Discovered

What if the patients' condition changes?

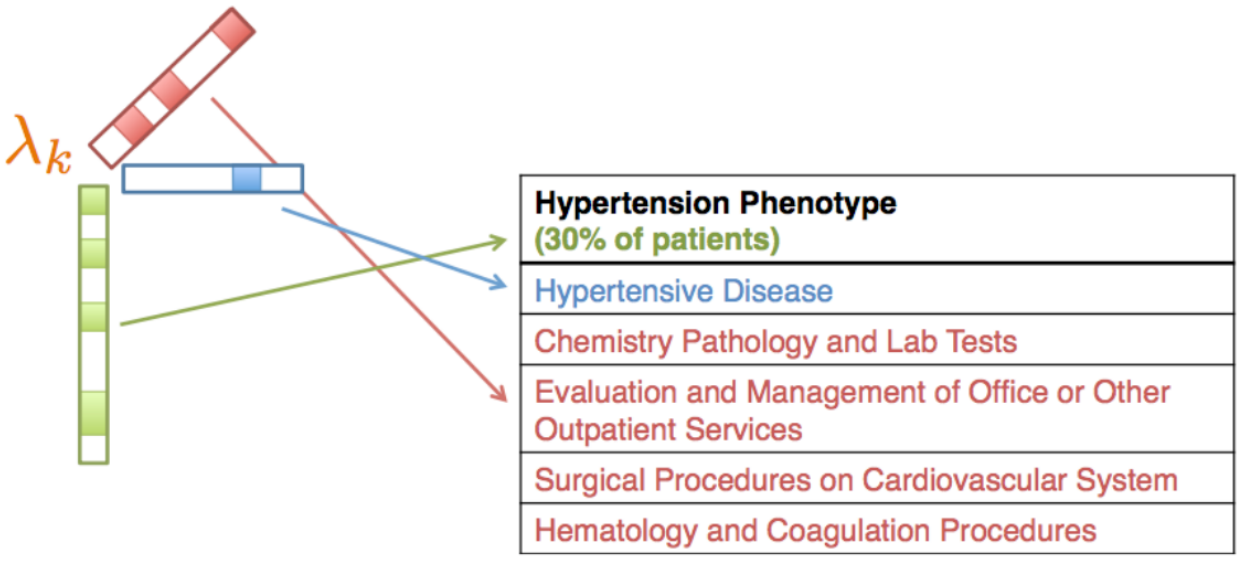


Figure: Phenotypes Discovered

Social Network Analysis

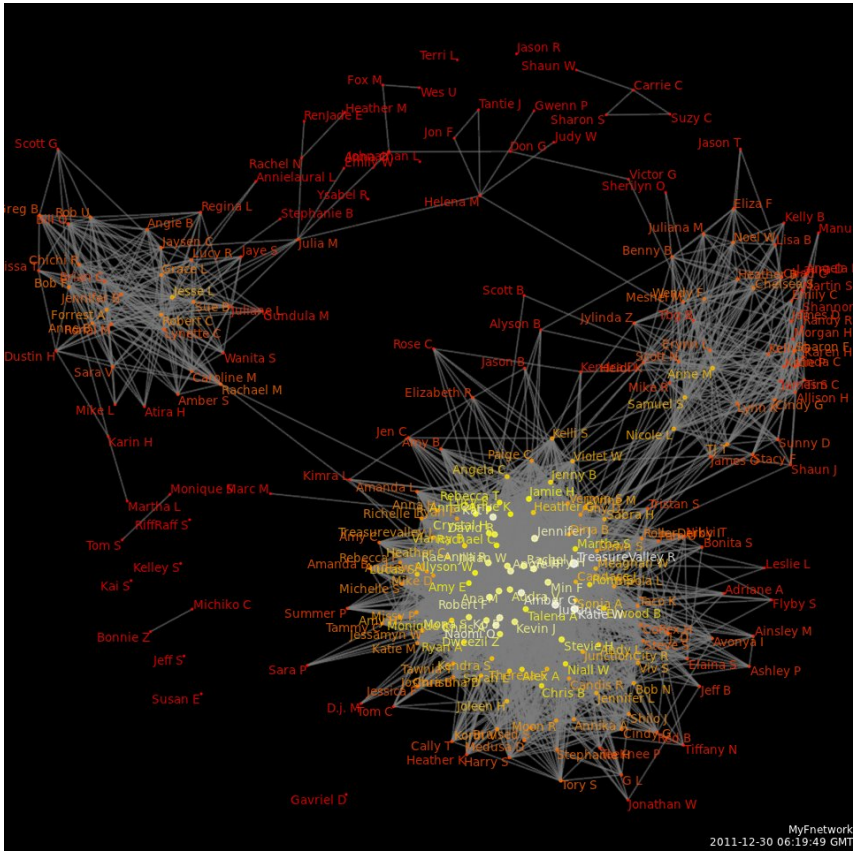


Figure: Social Network as a Graph

- Eigen Decomposition

What if the users' association changes with time?

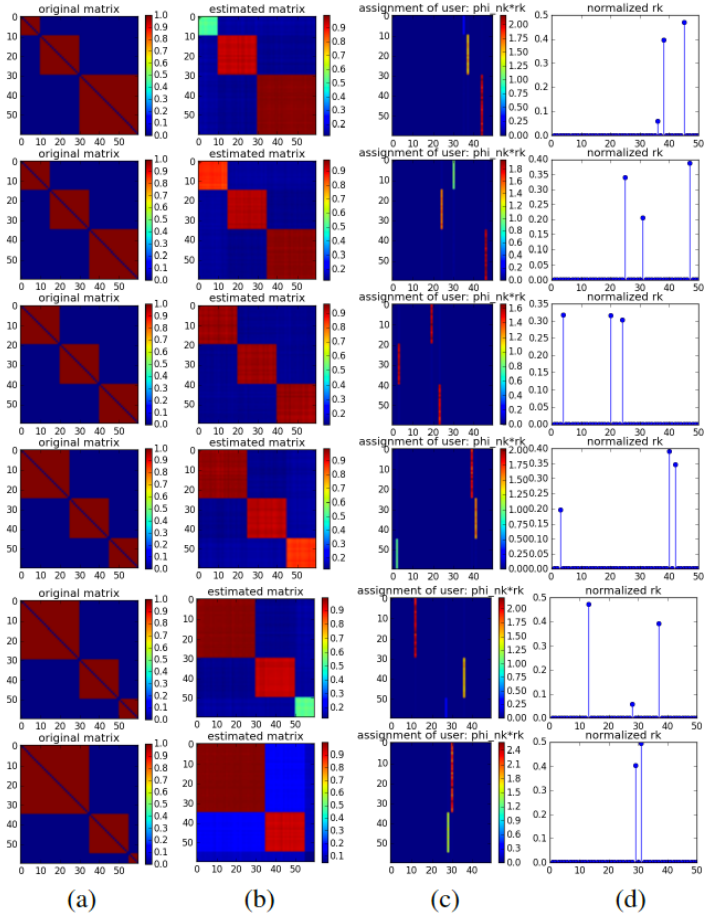


Figure: Static Modeling

What if the users' association changes with time?

- Eigen Decomposition

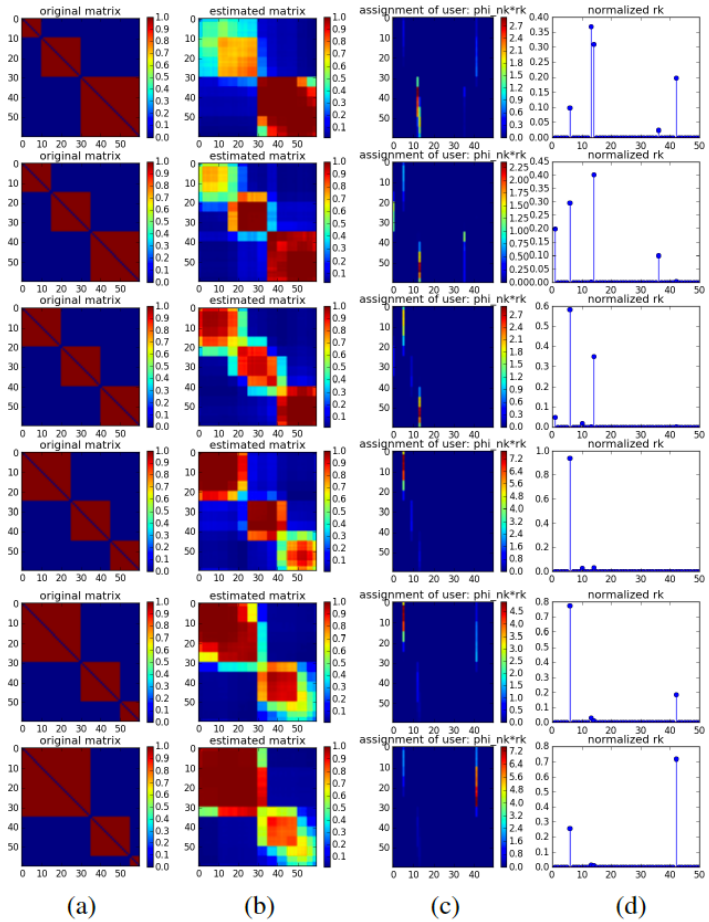


Figure: Dynamic Modeling

Knowledge Graph as Rank-reduced Tensor

- Significant reduction in storage
- Reduction in retrieval time/complexity

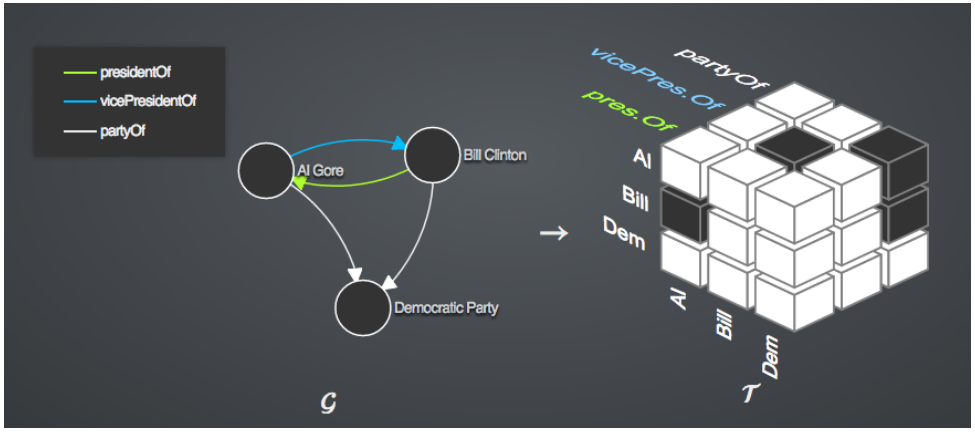


Figure: Knowledge Graph as Tensor

Conclusion

- Same mathematical tool used in solving problems in multiple domains
- Customization of a tool requires domain knowledge.
- Automated model selection is difficult.
- ML is not a magic, it requires a human to create one.

Questions?