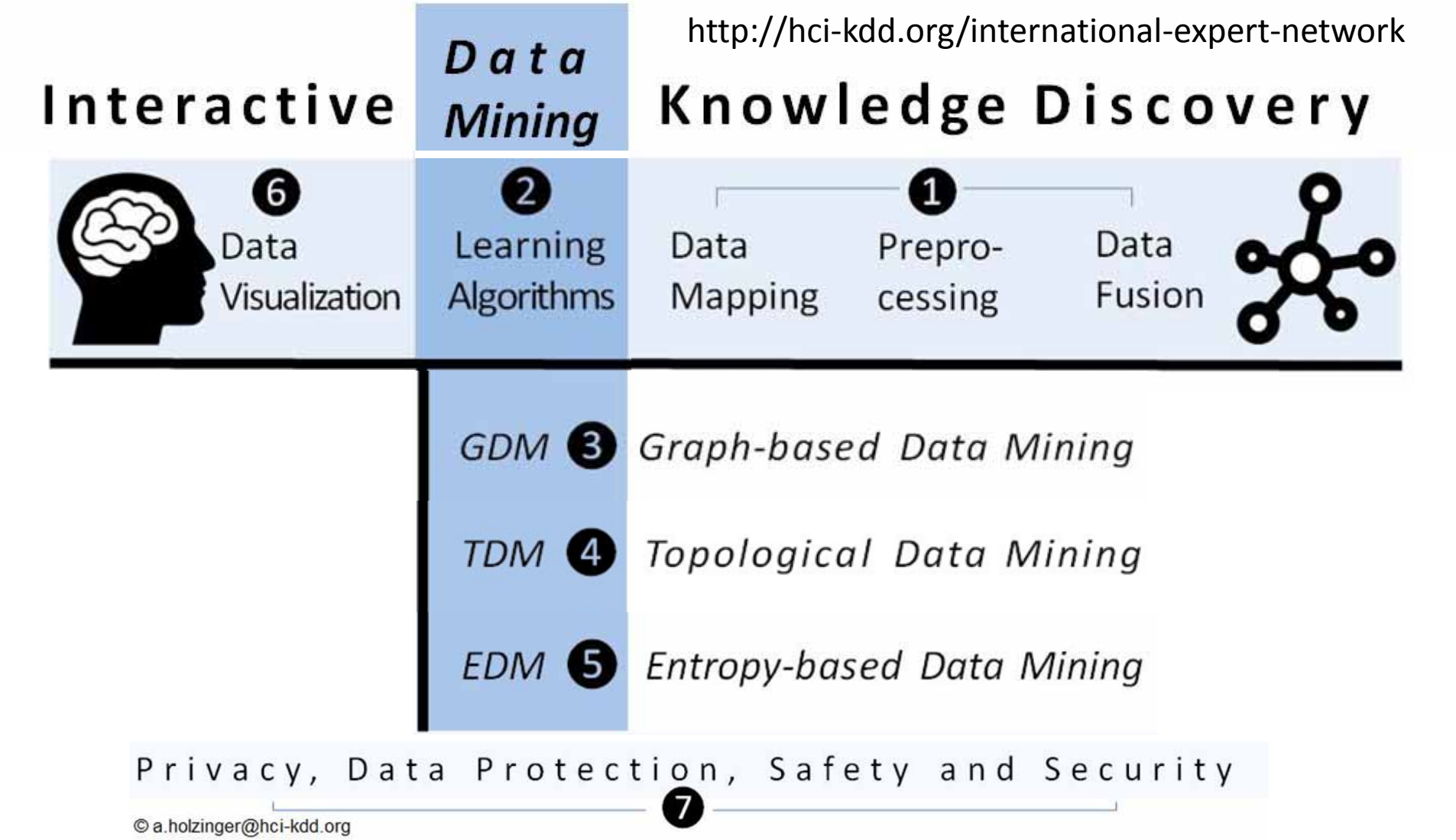
185.A83 Machine Learning for Health Informatics
2016S, VU, 2.0 h, 3.0 ECTS
Week 20 18.05.2016 17:00-20:00

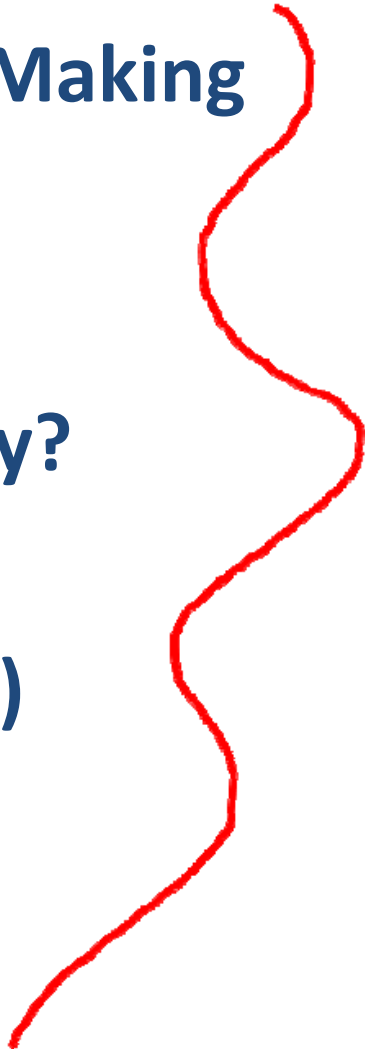
Probabilistic Graphical Models Part 2: From Bayesian Networks to Graph Bandits

a.holzinger@hci-kdd.org
http://hci-kdd.org/machine-learning-for-health-informatics-course

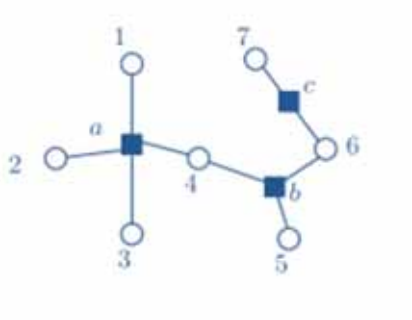


Holzinger, A. 2014. Trends in Interactive Knowledge Discovery for Personalized Medicine: Cognitive Science meets Machine Learning. IEEE Intelligent Informatics Bulletin, 15, (1), 6-14.

- 1) Graphical Models and Decision Making
- 2) Bayesian Networks
- 3) Machine Learning on Graphs
- 4) Little Excursus: What is similarity?
- 5) Probabilistic Topic Models
- 6) Graph Bandits (a very hot topic!)

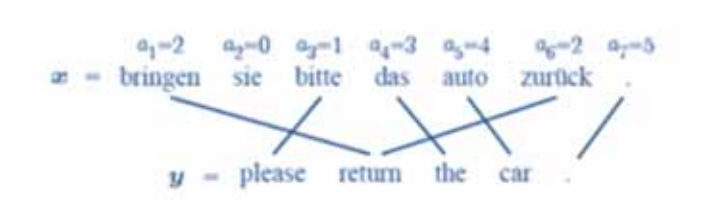


Let us start with a warm-up Quiz (solutions -> last page)



1

2



3

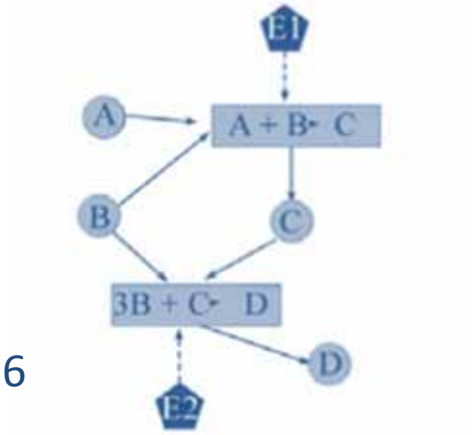
$$E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1, \ldots, x_n \mid d) U(x_1, \ldots, x_n, d)$$

4

h_1 = The identity of ORGANISM-1 is streptococcus
h_2 = PATIENT-1 is febrile
h_3 = The name of PATIENT-1 is John Jones

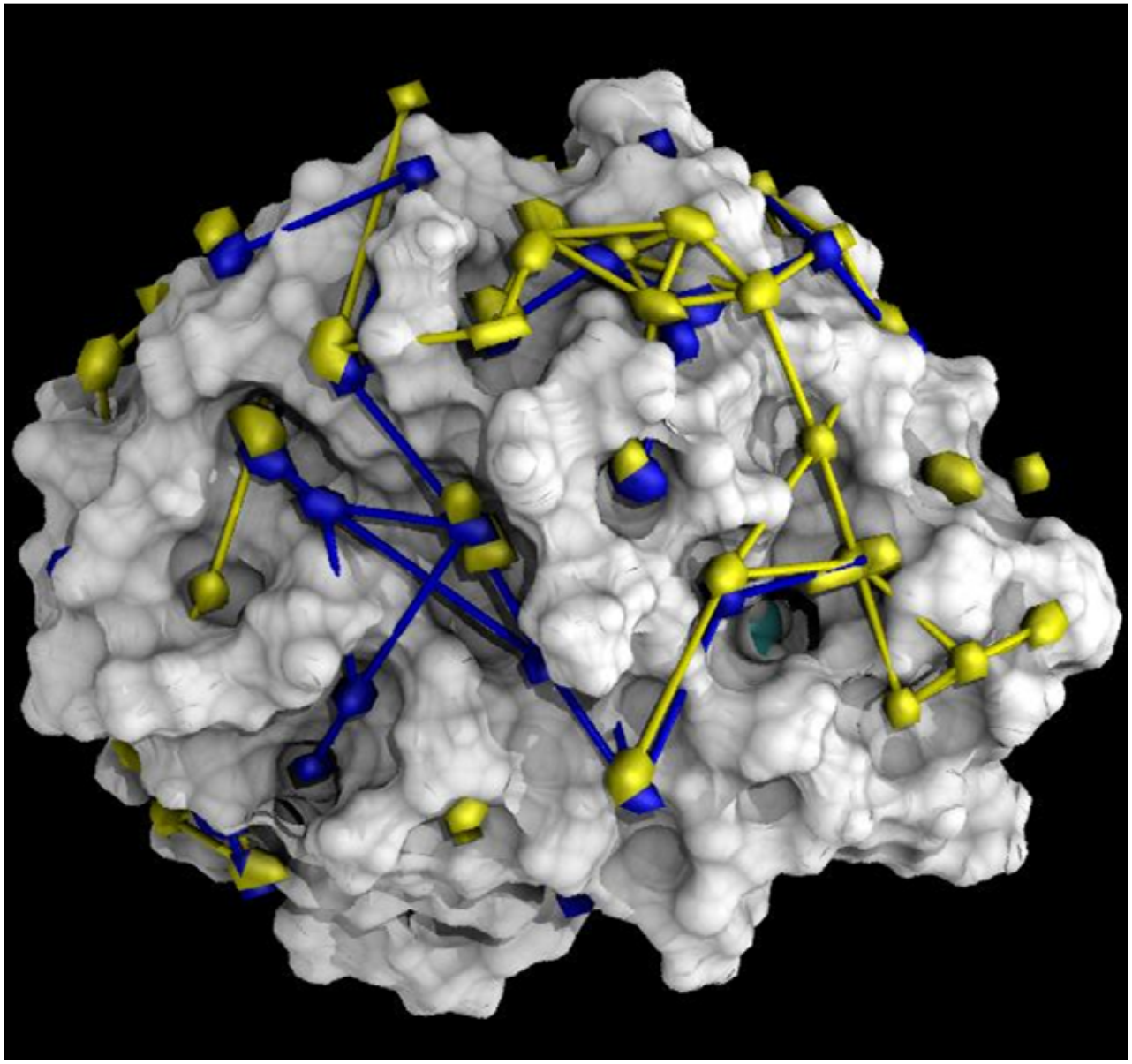
CF[h_1,E] = .8	:	There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus
CF[h_2,E] = −.3	:	There is weakly suggestive evidence (.3) that PATIENT-1 is not febrile
CF[h_3,E] = +1	:	It is definite (1) that the name of PATIENT-1 is John Jones

5



6

Quiz: last question …



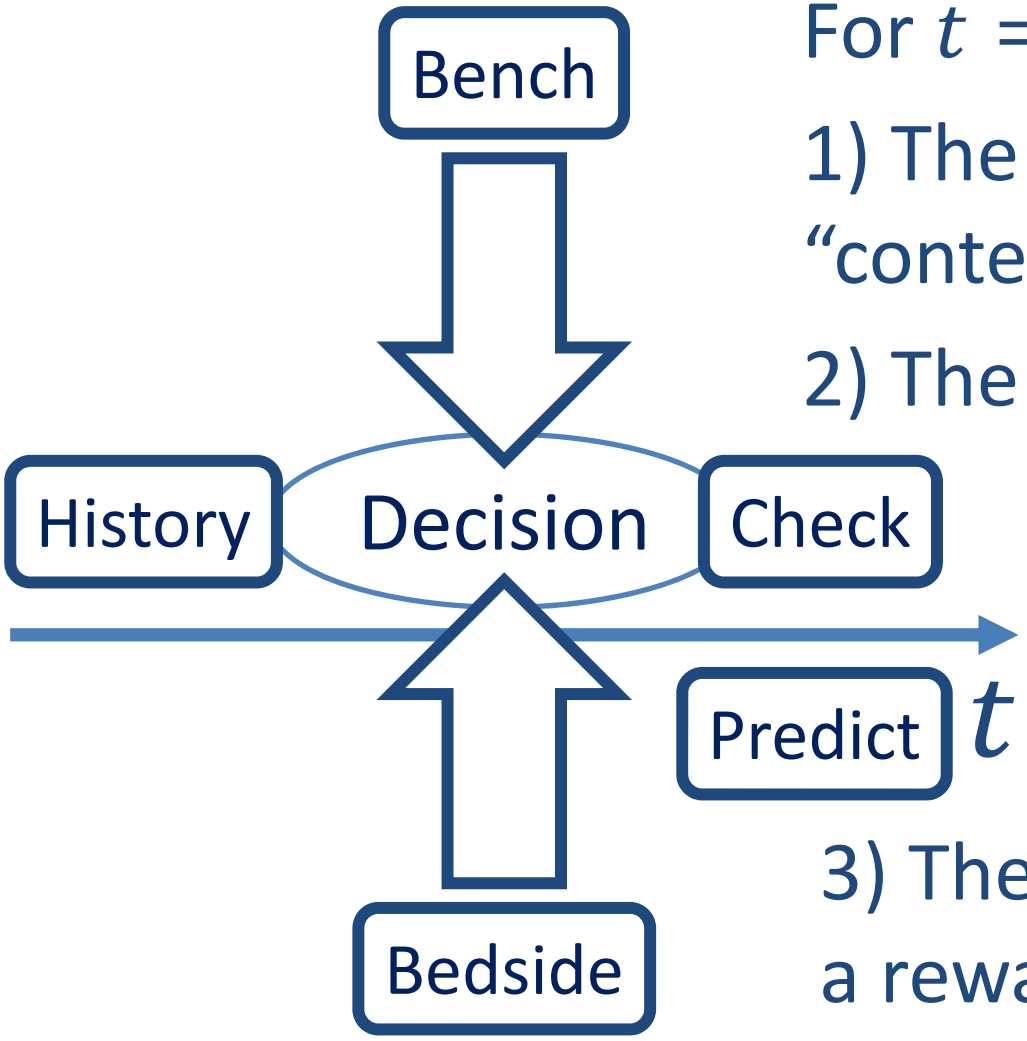
http://sbcb.bioch.ox.ac.uk/users/oliver/software/

1) Graphical Models and Decision Making

$\mathcal{M}$

$$\mathcal{D} \equiv \{X_1^{(i)}, X_2^{(i)}, \ldots, X_m^{(i)}\}_{i=1}^{N}$$

Goal: Learn an **optimal policy** for selecting best actions within a given **context**



For $t = 1, \dots, T$

1) The world produces a "context" $x_t \in X$

2) The learner selects an action $a_t \in \{1, \dots, K\}$

3) The world reacts with a reward $r_t(a_t) \in [0,1]$

GM are amongst the most important ML developments

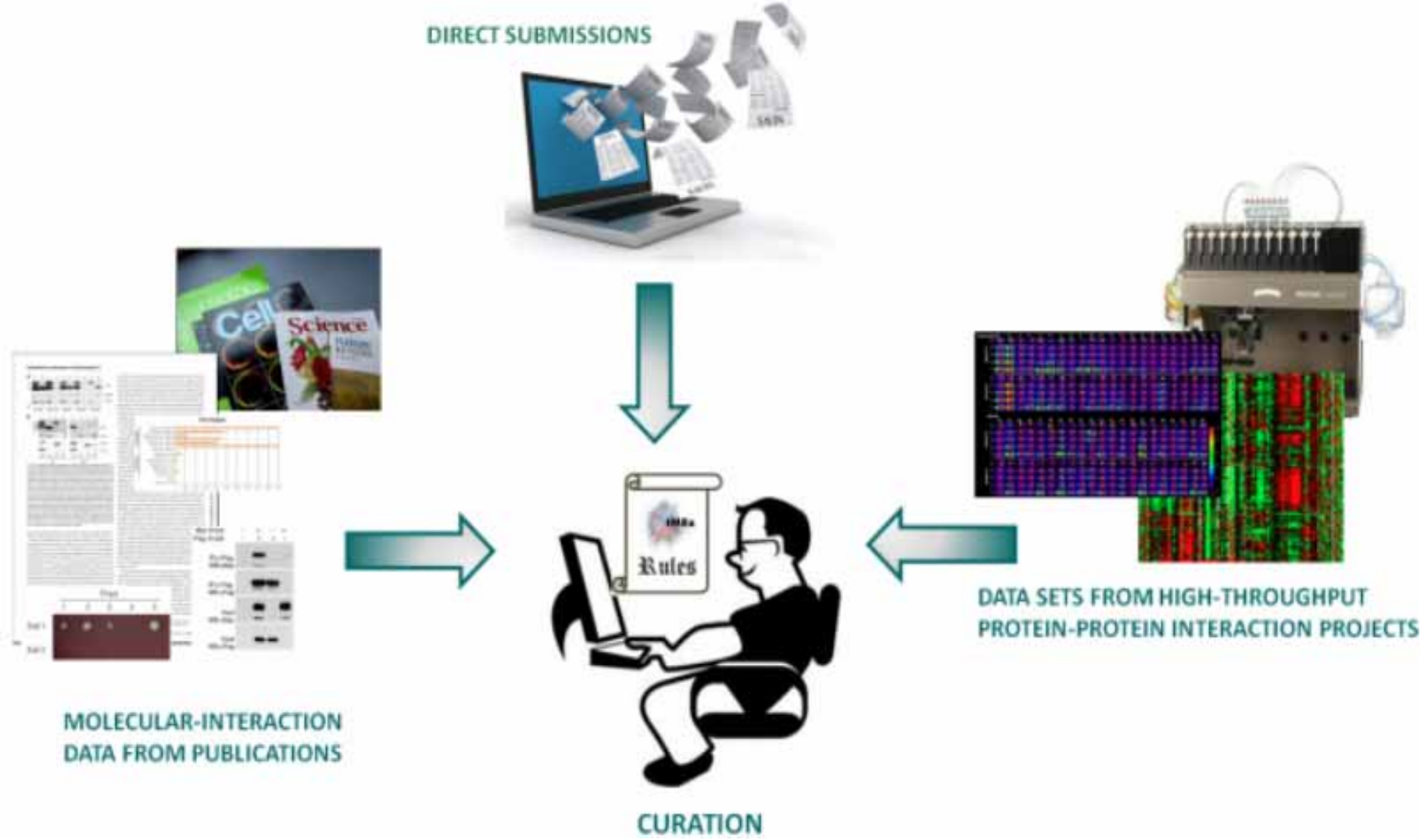
- Key Idea: Conditional independence assumptions are very useful – however: Naïve Bayes is extreme!
- X is *conditionally independent* of Y, given Z, if the P(X) governing X is independent of value Y, given value of Z:

$$(\forall i, j, k) P(X = x_i | Y = y_j, Z = z_k) = P(X = x_i | Z = z_k)$$

can be abbr. with $P(X|Y,Z) = P(X|Z)$

- Graphical models express sets of conditional independence assumptions via graph structure
- The graph structure plus associated parameters define joint probability distribution over the set of variables

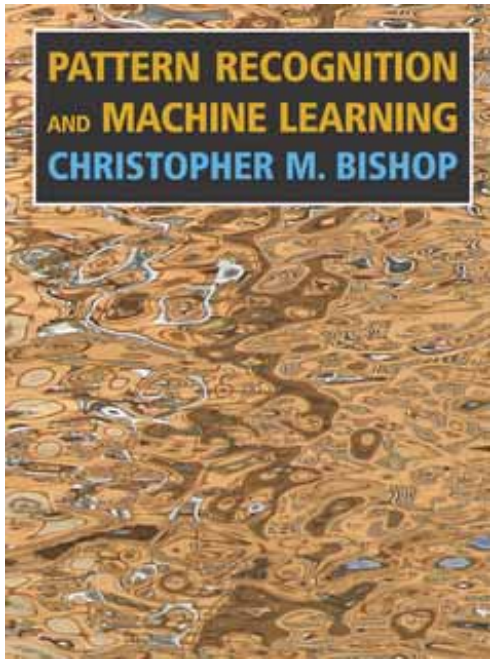
Where do the data come from?



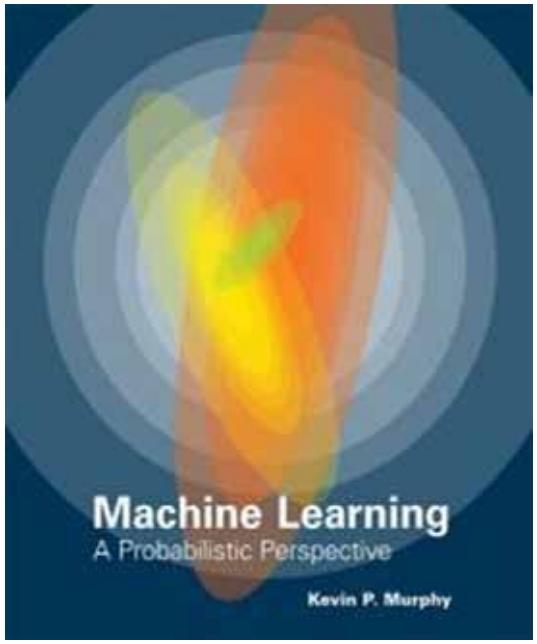
http://www.ebi.ac.uk/intact/

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties -> **probable information!**
- When we have big data but little knowledge automatic ML can help to gain insight:
- **Structure learning and prediction in large-scale biomedical networks with probabilistic graphical models**
- If we have little data and deal with NP-hard problems we still need the human-in-the-loop

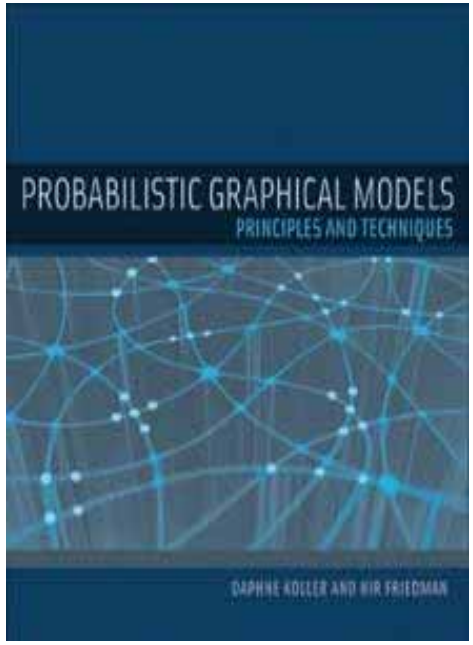
Basics and Background reading



Bishop, C. M. 2007. Pattern Recognition and Machine Learning, Heidelberg, Springer. Chapter 8 on graphical models openly available: http://research.microsoft.com/en-us/um/people/cmbishop/prml/

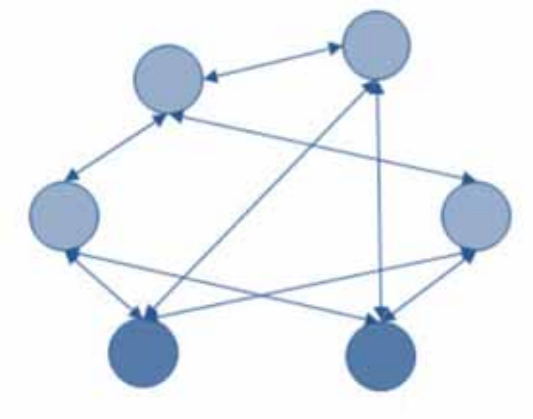


Murphy, K. P. 2012. Machine learning: a probabilistic perspective, MIT press. Chapter 26 (pp. 907) – Graphical model structure learning



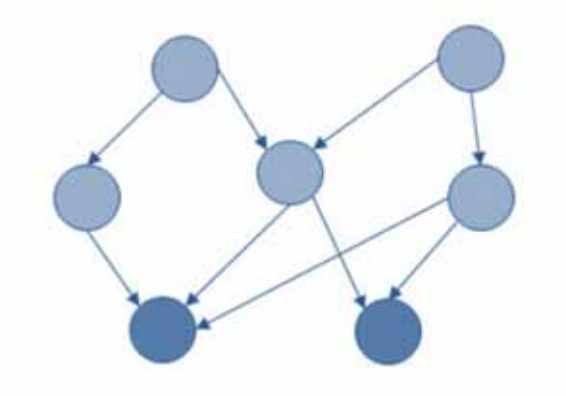
Koller, D. & Friedman, N. 2009. Probabilistic graphical models: principles and techniques, MIT press.

Three types of Probabilistic Graphical Models



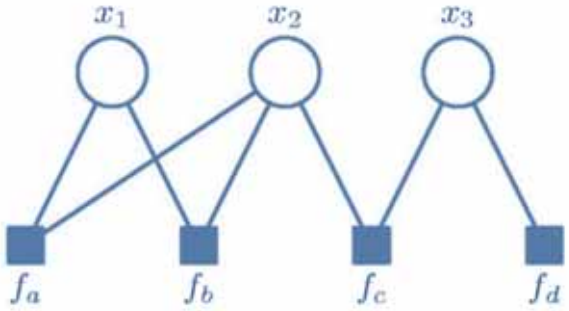
Undirected: Markov random fields, useful e.g. for computer vision (Details: Murphy 19)

$$P(\boldsymbol{X}) = \frac{1}{Z}\exp\left(\sum_{ij} W_{ij}\, x_i x_j + \sum_i x_i b_i\right)$$



Directed: Bayes Nets, useful for designing models (Details: Murphy 10)

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$



Factored: useful for inference/learning

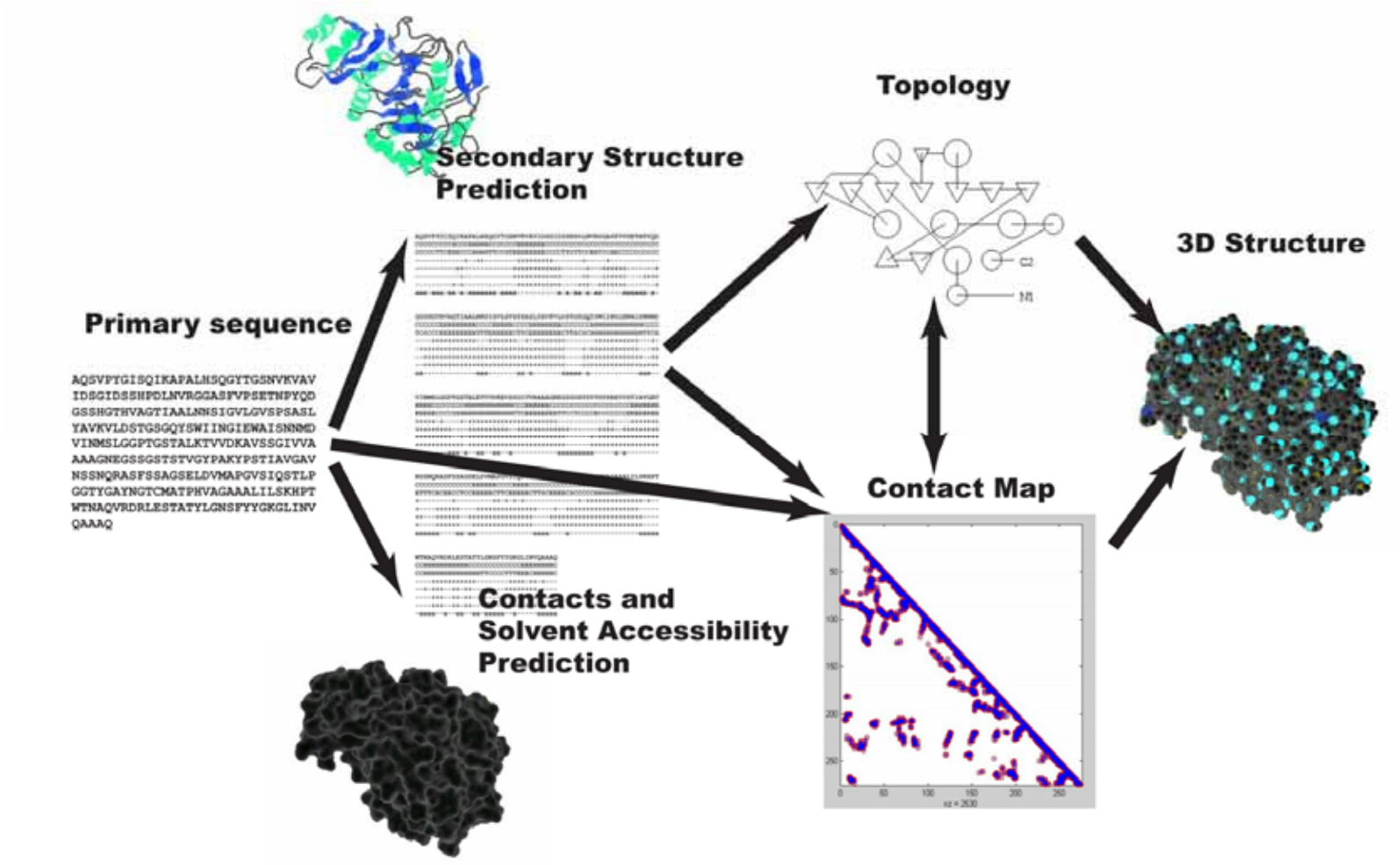
$$p(\mathbf{x}) = \prod_s f_s(\mathbf{x}_s)$$

- What is the advantage of factor graphs?

	Dependency	Efficient Inference	Usage
Bayesian Networks	Yes	Somewhat	Ancestral Generative Process
Markov Networks	Yes	No	Local Couplings and Potentials
Factor Graphs	No	Yes	Efficient, distributed inference

Table credit to Ralf Herbrich, Amazon

From structure to function prediction



Baldi, P. & Pollastri, G. 2003. The principled design of large-scale recursive neural network architectures--dag-rnns and the protein structure prediction problem. The Journal of Machine Learning Research, 4, 575-602.

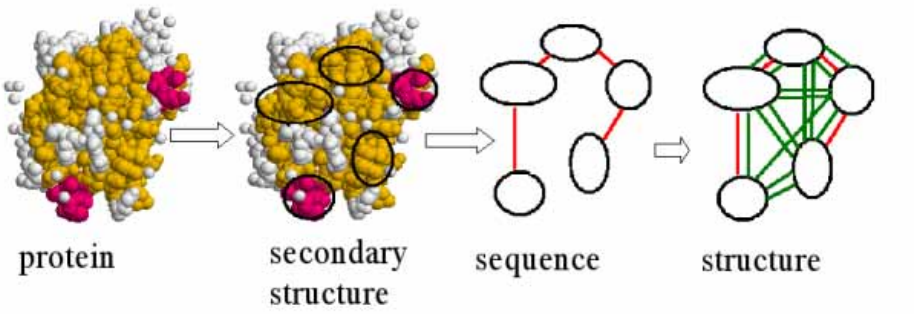
Protein Network Inference

- Hypothesis: most biological functions involve the interactions between many proteins, and the complexity of living systems arises as a result of such interactions.
- In this context, the problem of inferring a global protein network for a given organism,
- - using all (genomic) data of the organism,
- is one of the main challenges in computational biology

Yamanishi, Y., Vert, J.-P. & Kanehisa, M. 2004. Protein network inference from multiple genomic data: a supervised approach. Bioinformatics, 20, (suppl 1), i363-i370.

Problem: Is Graph Isomorphism NP-complete ?

Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S., Smola, A. J. & Kriegel, H.-P. 2005. Protein function prediction via graph kernels. Bioinformatics, 21, (suppl 1), i47-i56.



- Important for health informatics: Discovering relationships between biological components
- Unsolved problem in computer science:
- Can the graph isomorphism problem be solved in polynomial time?
  - So far, no polynomial time algorithm is known.
  - It is also not known if it is NP-complete
  - We know that subgraph-isomorphism is NP-complete

Example: Protein Network Inference

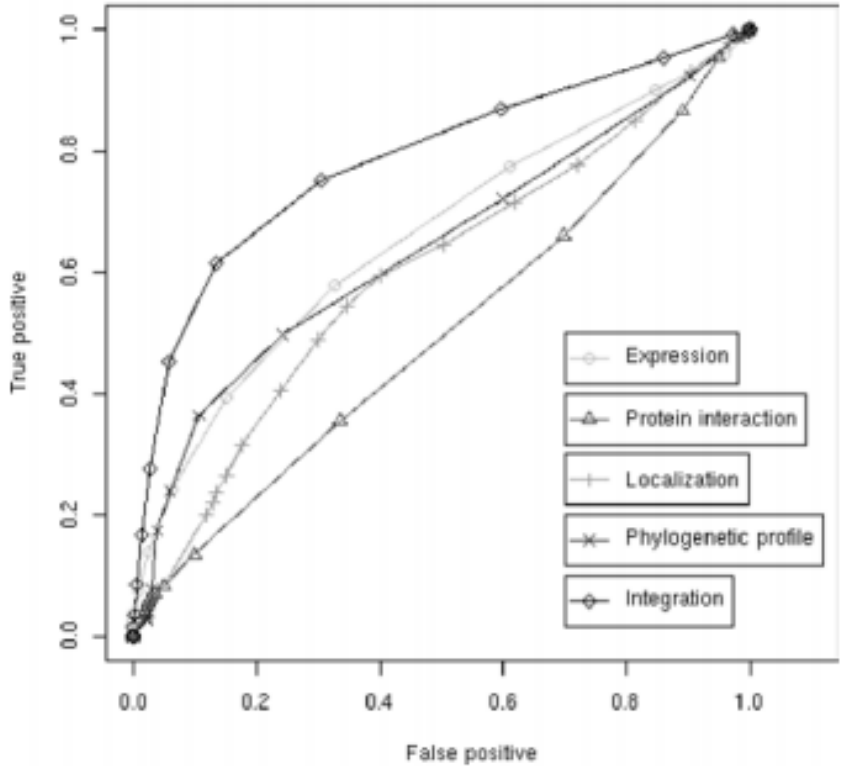
BIOINFORMATICS

Vol. 20 Suppl. 1 2004, pages i363–i370
DOI: 10.1093/bioinformatics/bth910

Protein network inference from multiple genomic data: a supervised approach

Y. Yamanishi[1,*], J.-P. Vert[2] and M. Kanehisa[1]

[1]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan and [2]Computational Biology group, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77305 Fontainebleau cedex, France



K_{exp} (Expression)
K_{ppi} (Protein interaction)
K_{loc} (Localization)
K_{phy} (Phylogenetic profile)
$K_{\text{exp}} + K_{\text{ppi}} + K_{\text{loc}} + K_{\text{phy}}$ (Integration)

Example: Data fusion and Protein Annotation

BIOINFORMATICS

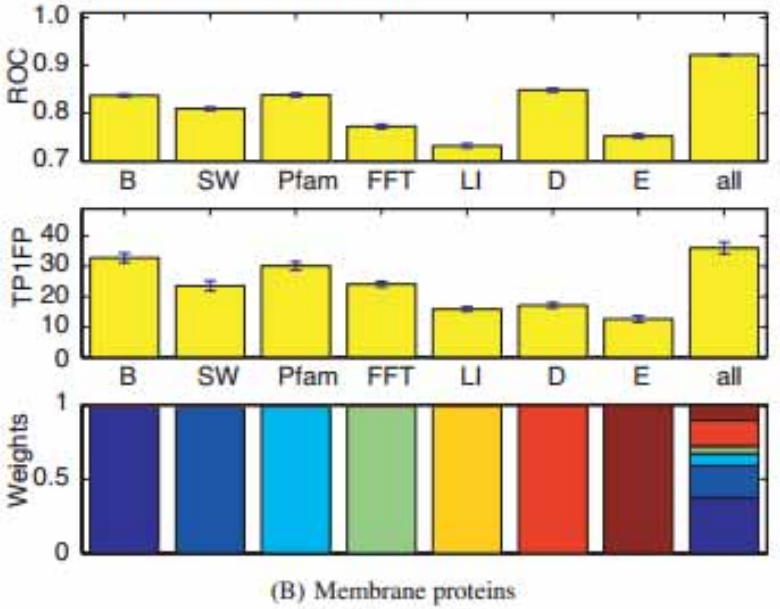
Vol. 20 no. 16 2004, pages 2626–2635
doi:10.1093/bioinformatics/bth294

A statistical framework for genomic data fusion

Gert R. G. Lanckriet[1], Tijl De Bie[3], Nello Cristianini[4], Michael I. Jordan[2] and William Stafford Noble[5,*]

[1]Department of Electrical Engineering and Computer Science, [2]Division of Computer Science, Department of Statistics, University of California, Berkeley 94720, USA, [3]Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven 3001, Belgium, [4]Department of Statistics, University of California, Davis 95618, USA and [5]Department of Genome Sciences, University of Washington, Seattle 98195, USA

Kernel	Data	Similarity measure
K_{SW}	protein sequences	Smith-Waterman
K_{B}	protein sequences	BLAST
K_{Pfam}	protein sequences	Pfam HMM
K_{FFT}	hydropathy profile	FFT
K_{LI}	protein interactions	linear kernel
K_{D}	protein interactions	diffusion kernel
K_{E}	gene expression	radial basis kernel
K_{RND}	random numbers	linear kernel



(B) Membrane proteins

Lanckriet, G. R., De Bie, T., Cristianini, N., Jordan, M. I. & Noble, W. S. 2004. A statistical framework for genomic data fusion. Bioinformatics, 20, (16), 2626-2635.

2) Bayesian Networks "Bayes' Nets"

Bayesian Network (BN) - Definition

- is a **probabilistic model**, consisting of two parts:
- 1) a dependency structure and
- 2) local probability models.

$$p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i \mid Pa(x_i))$$

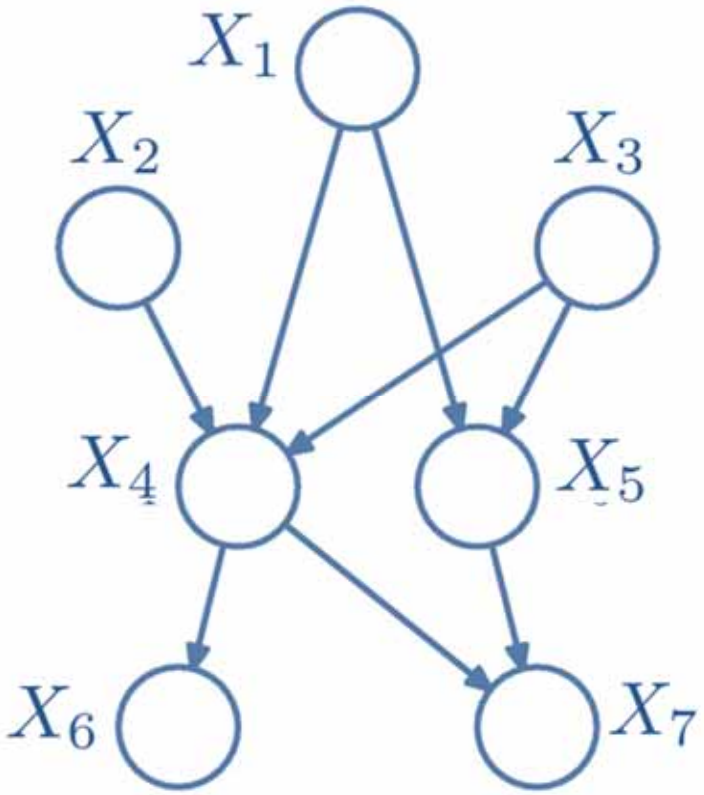
Where $Pa(x_i)$ are the parents of x_i

BN inherently model the **uncertainty in the data.** They are a successful marriage between probability theory and graph theory; allow to model a multidimensional probability distribution in a sparse way by searching independency relations in the data. Furthermore this model allows different strategies to integrate two data sources.

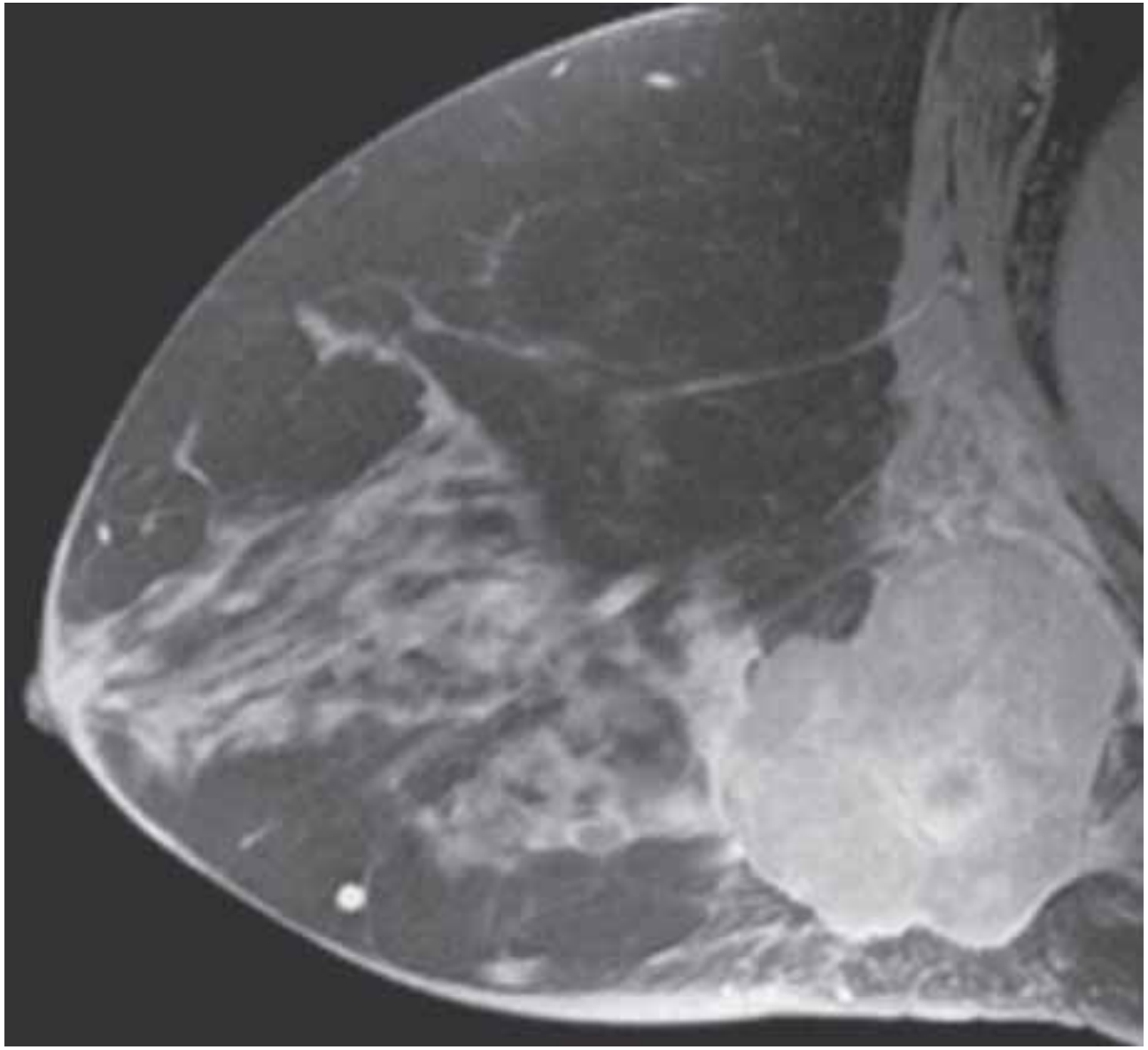
Pearl, J. (1988) *Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, Morgan Kaufmann.*

Example: Directed Bayesian Network with 7 nodes

$$p(X_1, \ldots, X_7) =$$
$$p(X_1)p(X_2)p(X_3)p(X_4|X_1, X_2, X_3) \cdot$$
$$p(X_5|X_1, X_3)p(X_6|X_4)p(X_7|X_4, X_5)$$



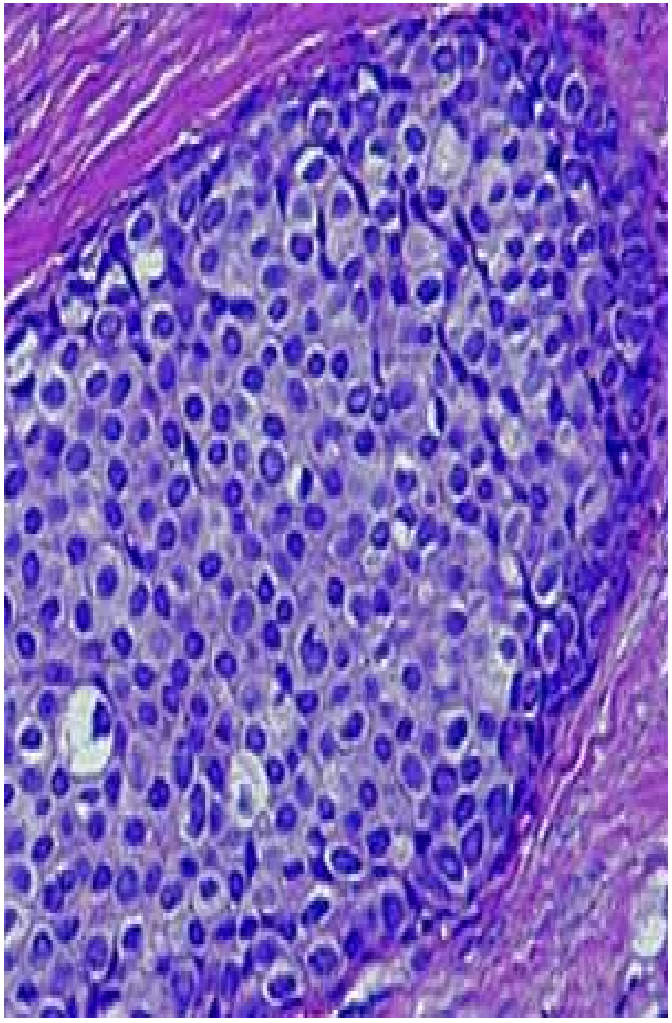
Clinical Case Example



Overmoyer, B. A., Lee, J. M. & Lerwill, M. F. (2011) Case 17-2011 A 49-Year-Old Woman with a Mass in the Breast and Overlying Skin Changes. *New England Journal of Medicine, 364, 23, 2246-2254.*

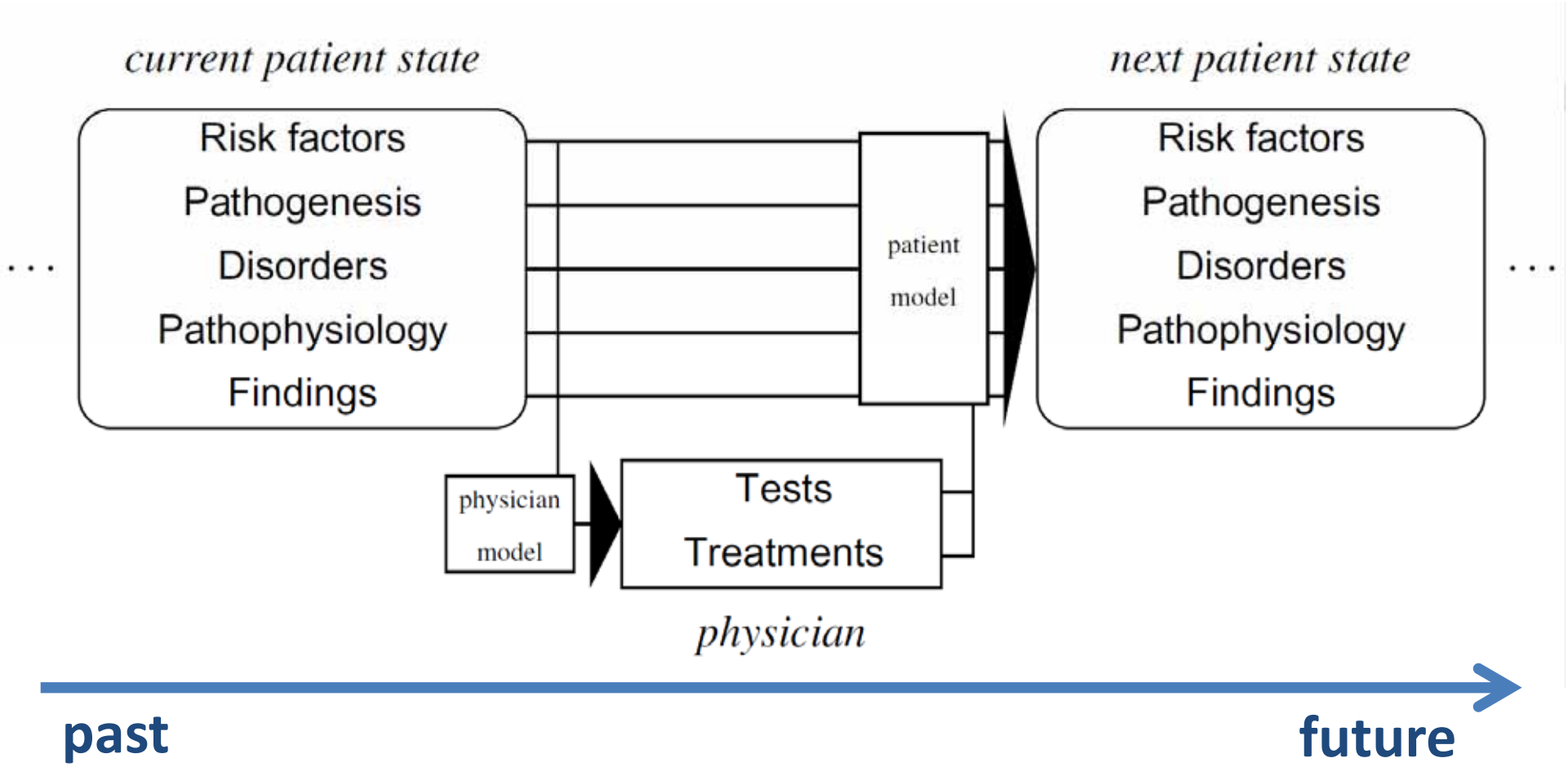
Important in Clinical practice -> prognosis !

- = the prediction of the future course of a disease conditional on the patient's history and a projected treatment strategy
- Danger: probable Information !
- Therefore valid prognostic models can be of great benefit for clinical decision making and of great value to the patient, e.g., for notification and quality of-life decisions



Knaus, W. A., Wagner, D. P. & Lynn, J. (1991) Short-term mortality predictions for critically ill hospitalized adults: science and ethics. *Science, 254, 5030, 389.*

Predicting the future on past data and present status



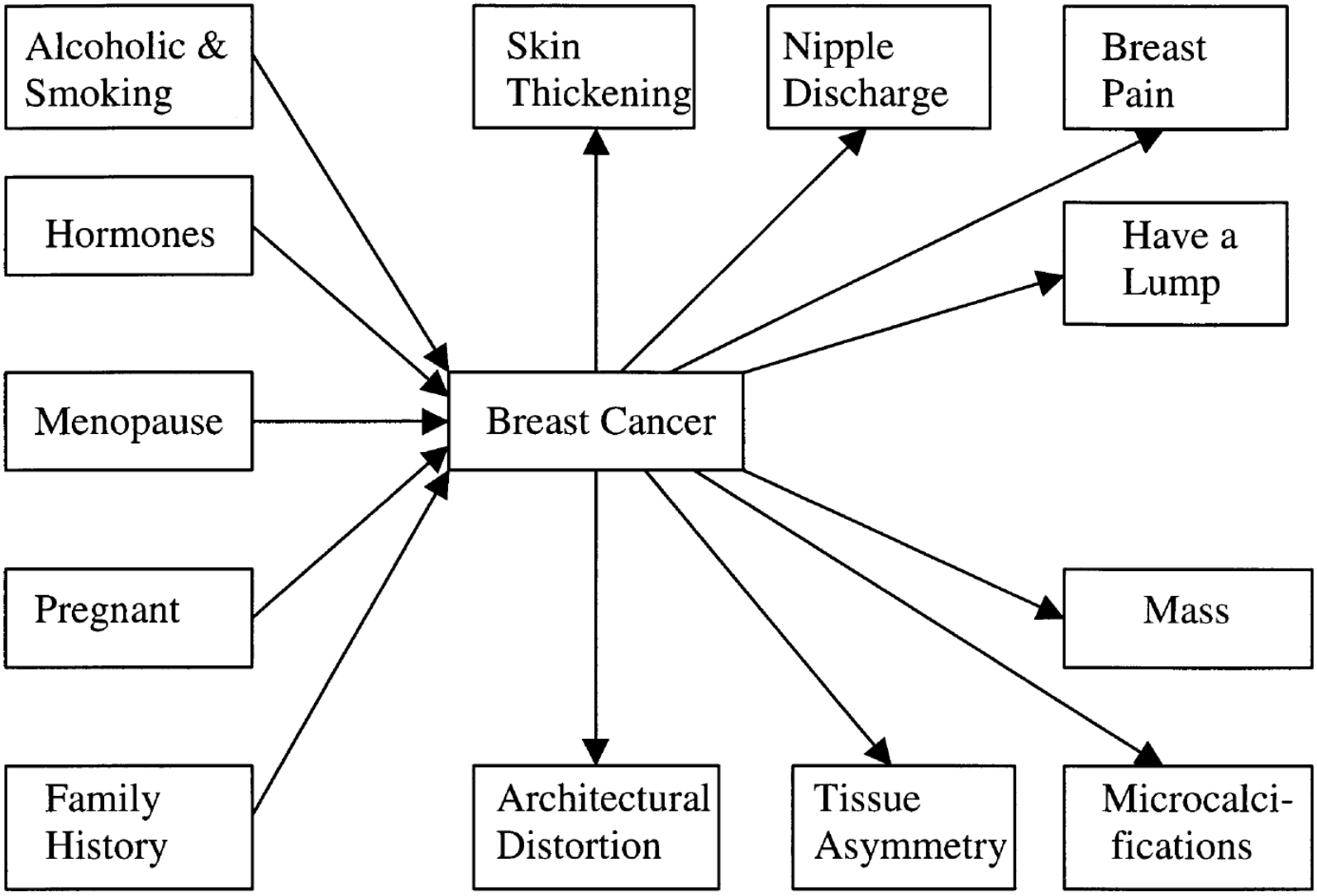
van Gerven, M. A. J., Taal, B. G. & Lucas, P. J. F. (2008) Dynamic Bayesian networks as prognostic models for clinical patient management. *Journal of Biomedical Informatics, 41, 4, 515-529.*

Example: Breast cancer - Probability Table

Category	Node description	State description
Diagnosis	Breast cancer	Present, absent.
Clinical history	Habit of drinking alcoholic beverages and smoking	Yes, no.
	Taking female hormones	Yes, no.
	Have gone through menopause	Yes, no.
	Have ever been pregnant	Yes, no.
	Family member has breast cancer	Yes, no.
Physical findings	Nipple discharge	Yes, no.
	Skin thickening	Yes, no.
	Breast pain	Yes, no.
	Have a lump(s)	Yes, no.
Mammographic findings	Architectural distortion	Present, absent.
	Mass	Score from one to three, score from four to five, absent
	Microcalcification cluster	Score from one to three, score from four to five, absent
	Asymmetry	Present, absent.

Wang, X. H., et al. (1999) Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics, 54, 2, 115-126.*

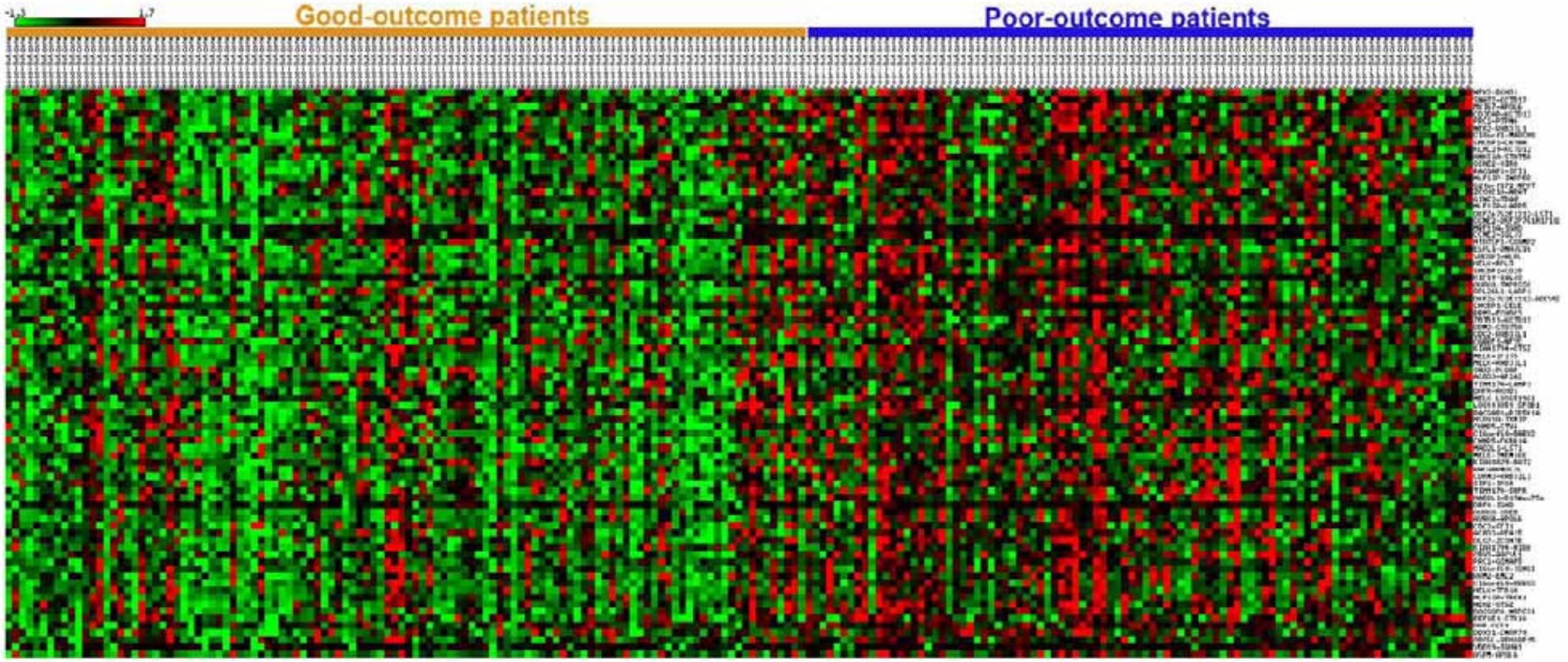
Breast cancer – big picture – state of 1999



Wang, X. H., et al. (1999) Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *International Journal of Medical Informatics, 54, 2, 115-126.*

10 years later: Integration of microarray data

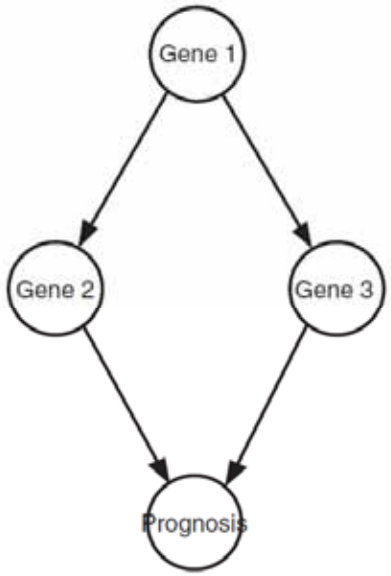
- Integrating microarray data from multiple studies to increase sample size;
- = approach to the development of more robust prognostic tests



Xu, L., Tan, A., Winslow, R. & Geman, D. (2008) Merging microarray data from separate breast cancer studies provides a robust prognostic test. *BMC Bioinformatics, 9, 1, 125-139.*

Example: BN with four binary variables

Gene 1	
P(on)	0.8
P(off)	0.2



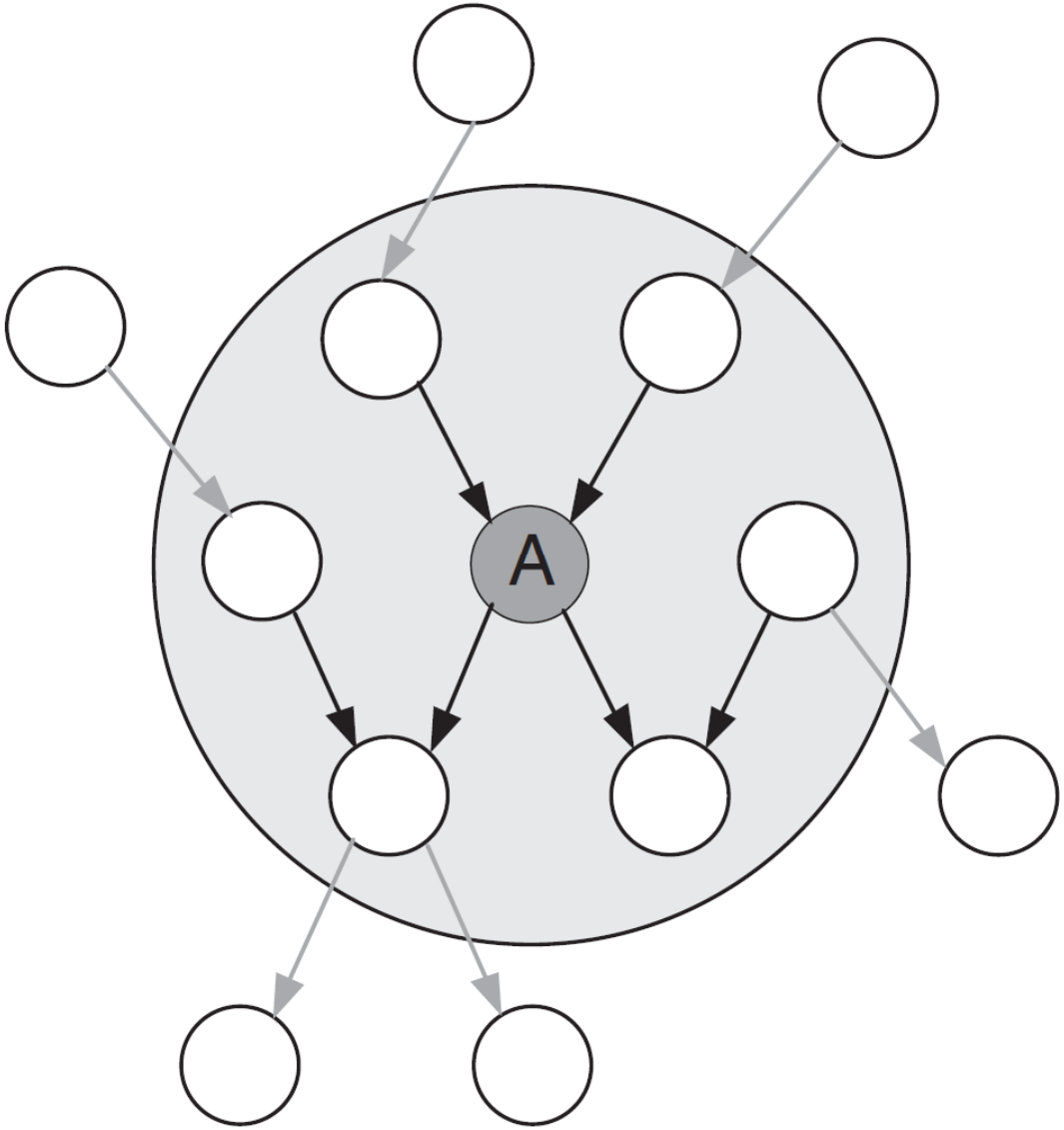
Gene 2	Gene 1 on	Gene 1 off
P(on)	0.3	0.6
P(off)	0.7	0.4

Gene 2	Gene 1 on	Gene 1 off
P(on)	0.3	0.6
P(off)	0.7	0.4

Prognosis	Gene 2 on Gene 3 on	Gene 2 on Gene 3 off	Gene 2 off Gene 2 on	Gene 2 off Gene 3 off
P(good)	0.6	0.1	0.9	0.5
P(poor)	0.4	0.9	0.1	0.5

Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*

Concept Markov-Blanket



Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*

Dependency Structure -> first step (1/2)

- First the structure is learned using a search strategy.
- Since the number of possible structures increases super exponentially with the number of variables,
- the well-known greedy search algorithm K2 can be used in combination with the Bayesian Dirichlet (BD) scoring metric:

$$p(S|D) \propto p(S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \left[\frac{\Gamma(N'_{ij})}{\Gamma(N'_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N'_{ijk} + N_{ijk})}{\Gamma(N'_{ijk})} \right]$$

N_{ijk} … number of cases in the data set D
having variable i in state k associated with the j-th instantiation of its parents in current structure S.
n is the total number of variables.

Dependency Structure – first step (2/2)

- Next, N_{ij} is calculated by summing over all states of a variable:
- $N_{ij} = \sum_{k=1}^{r_i} N_{ijk} \cdot N'_{ijk}$ and N'_{ij} have similar meanings but refer to prior knowledge for the parameters.
- When no knowledge is available they are estimated using $N_{ijk} = N/(r_i q_i)$
- with N the equivalent sample size,
- r_i the number of states of variable i and
- q_i the number of instantiations of the parents of variable i.
- $\Gamma(.)$ corresponds to the gamma distribution.
- Finally $p(S)$ is the prior probability of the structure.
- $p(S)$ is calculated by:
- $p(S) = \prod_{i=1}^{n} \prod_{l_i=1}^{p_i} p(l_i \longrightarrow x_i) \prod_{m_i=1}^{o_i} p(m_i x_i)$
- with p_i the number of parents of variable x_i and o_i all the variables that are not a parent of x_i.
- Next, $p(a \longrightarrow b)$ is the probability that there is an edge from a to b while $p(ab)$ is the inverse, i.e. the probability that there is no edge from a to b

Parameter learning -> second step

- Estimating the parameters of the local probability models corresponding with the dependency structure.
- CPTs are used to model these local probability models.
- For each variable and instantiation of its parents there exists a CPT that consists of a set of parameters.
- Each set of parameters was given a uniform Dirichlet prior:

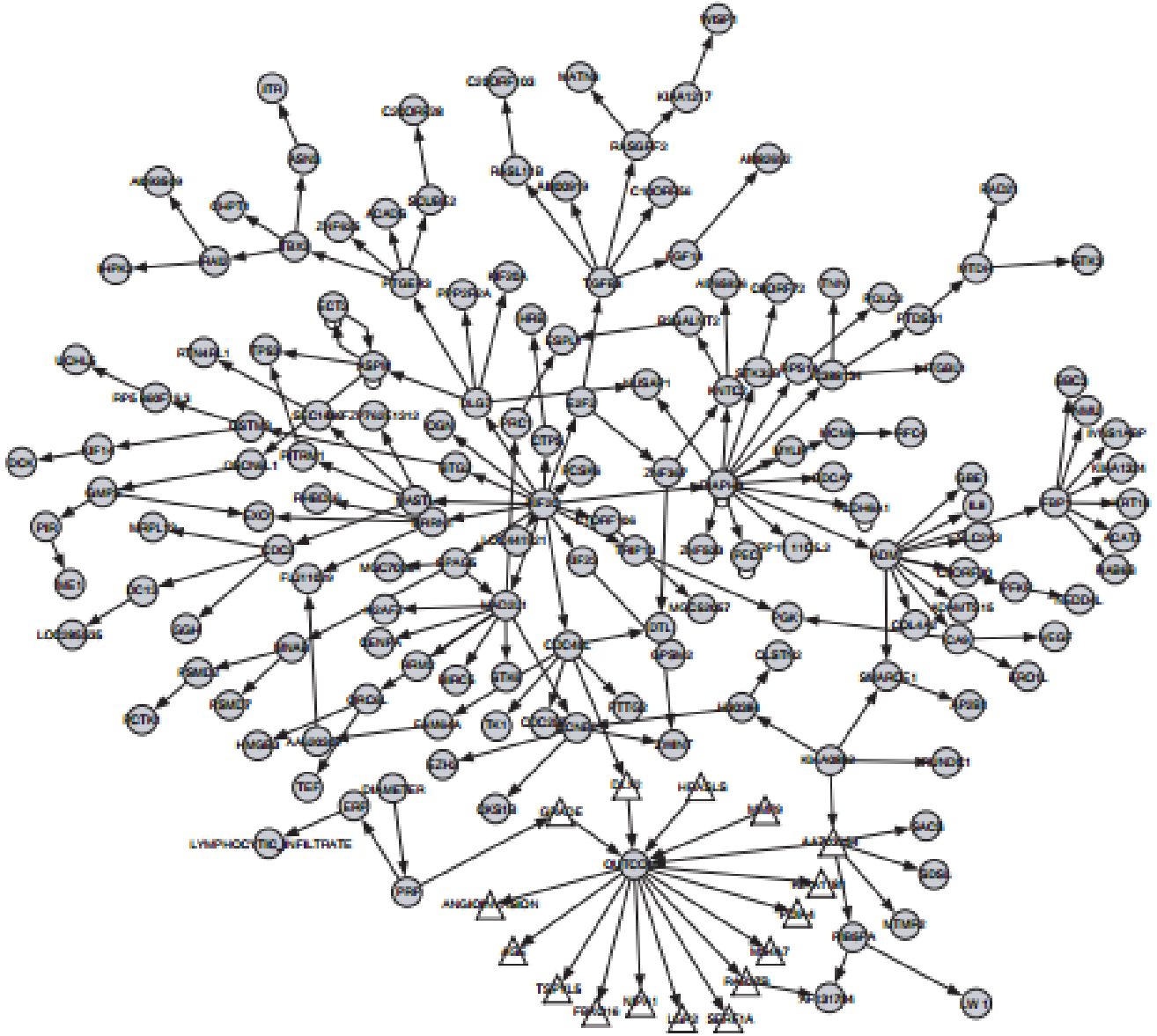
$$\mathrm{p}(\theta_{ij}|S) = Dir(\theta_{ij}|N'_{ij1}, \dots, N'_{ijk}, \dots, N'_{ijr_i})$$

Note: With θ_{ij} a parameter set where i refers to the variable and j to the j-th instantiation of the parents in the current structure. θ_{ij} contains a probability for every value of the variable x_i given the current instantiation of the parents. Dir corresponds to the Dirichlet distribution with $(N'_{ij1}, \dots, N'_{ijr_i})$ as parameters of this Dirichlet distribution. Parameter learning then consists of updating these Dirichlet priors with data. This is straightforward because the multinomial distribution that is used to model the data, and the Dirichlet distribution that models the prior, are conjugate distributions. This results in a Dirichlet posterior over the parameter set:

$$\mathrm{p}(\theta_{ij}|D,S) = Dir(\theta_{ij}|N'_{ij1} + N_{ij1}, \dots, N'_{ijk} + N_{ijk}, \dots, N'_{ijr_i} + N_{ijr_i})$$

with N_{ijk} defined as before.

Predicting the prognosis of breast cancer (integrated a.)



Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics, 22, 14, 184-190.*

Inference in Bayes Nets is intractable (NP-complete!)

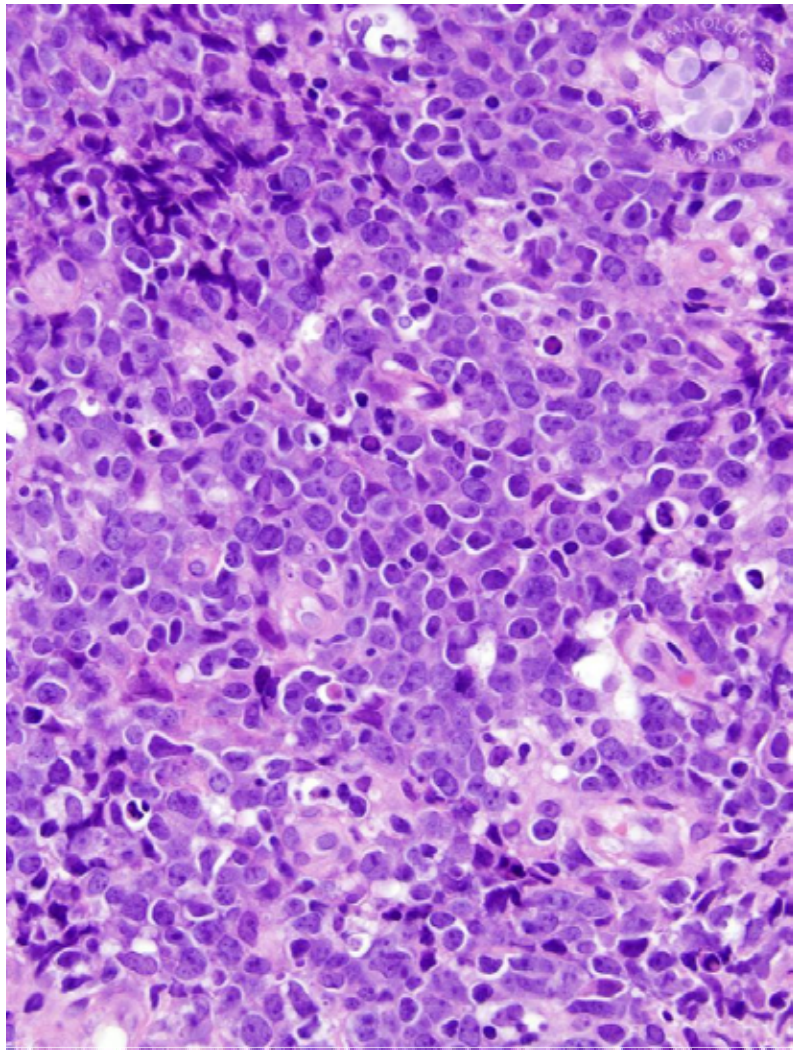
- For certain cases it is tractable if:
  - Just one variable is unobserved
  - We have singly connected graphs (no undirected loops -> belief propagation)
  - Assigning probability to fully observed set of variables
- Possibility: Monte Carlo Methods (generate many samples according to the Bayes Net distribution and then count the results)
- Otherwise: approximate solutions, NOTE:

Sometimes it is better to have an approximate solution to a complex problem – than a perfect solution to a simplified problem

3) Machine Learning on Graphs

Example: Lymphoma is the most common blood cancer

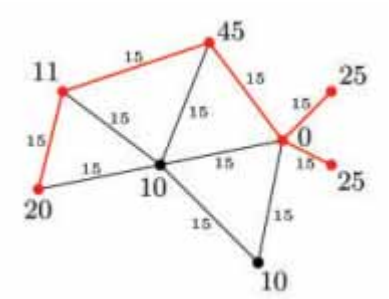
The two main forms of lymphoma are Hodgkin lymphoma and non-Hodgkin lymphoma (NHL). Lymphoma occurs when cells of the immune system called lymphocytes, a type of white blood cell, grow and multiply uncontrollably. Cancerous lymphocytes can travel to many parts of the body, including the lymph nodes, spleen, bone marrow, blood, or other organs, and form a mass called a tumor. The body has two main types of lymphocytes that can develop into lymphomas: B-lymphocytes (B-cells) and T-lymphocytes (T-cells).

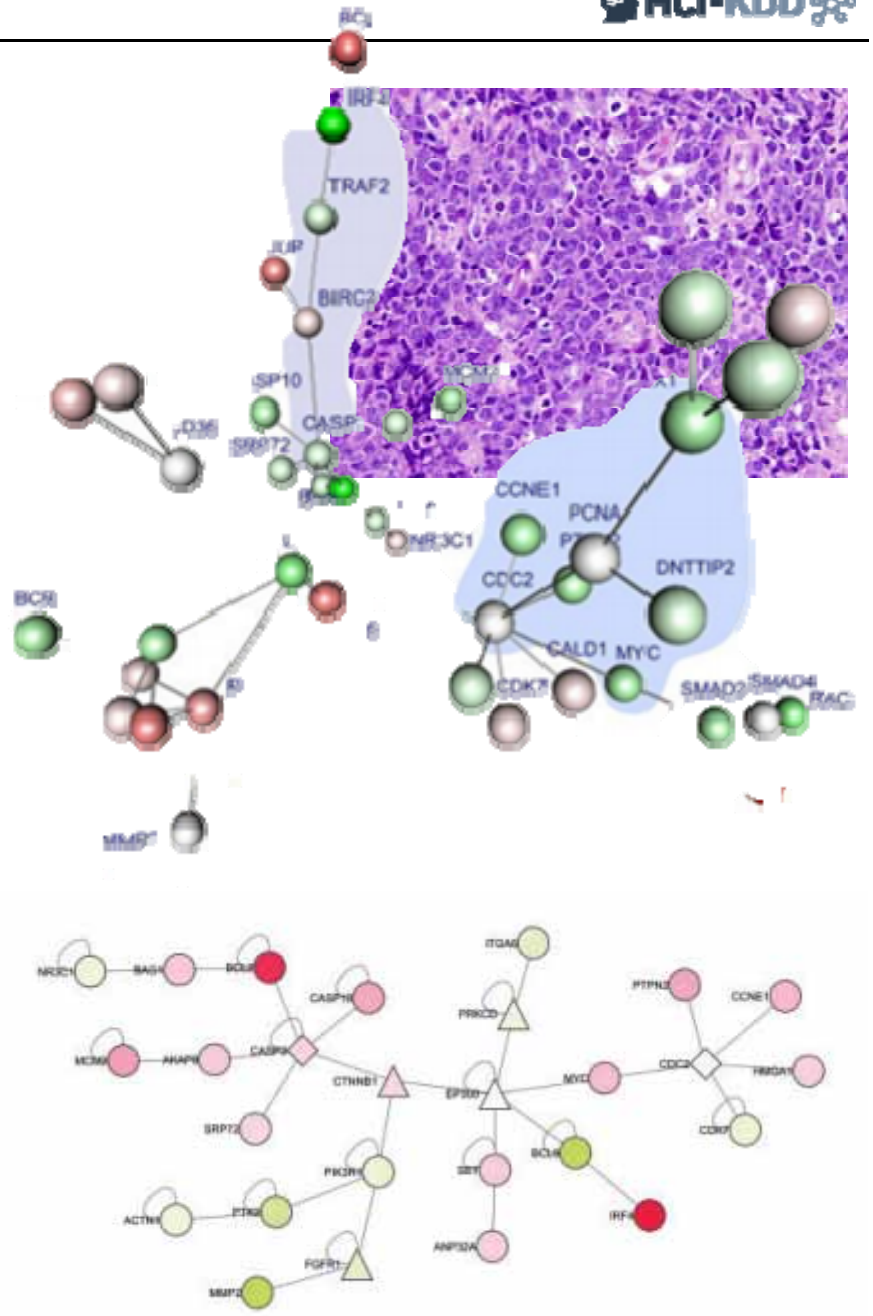


www.lymphoma.org

http://imagebank.hematology.org/

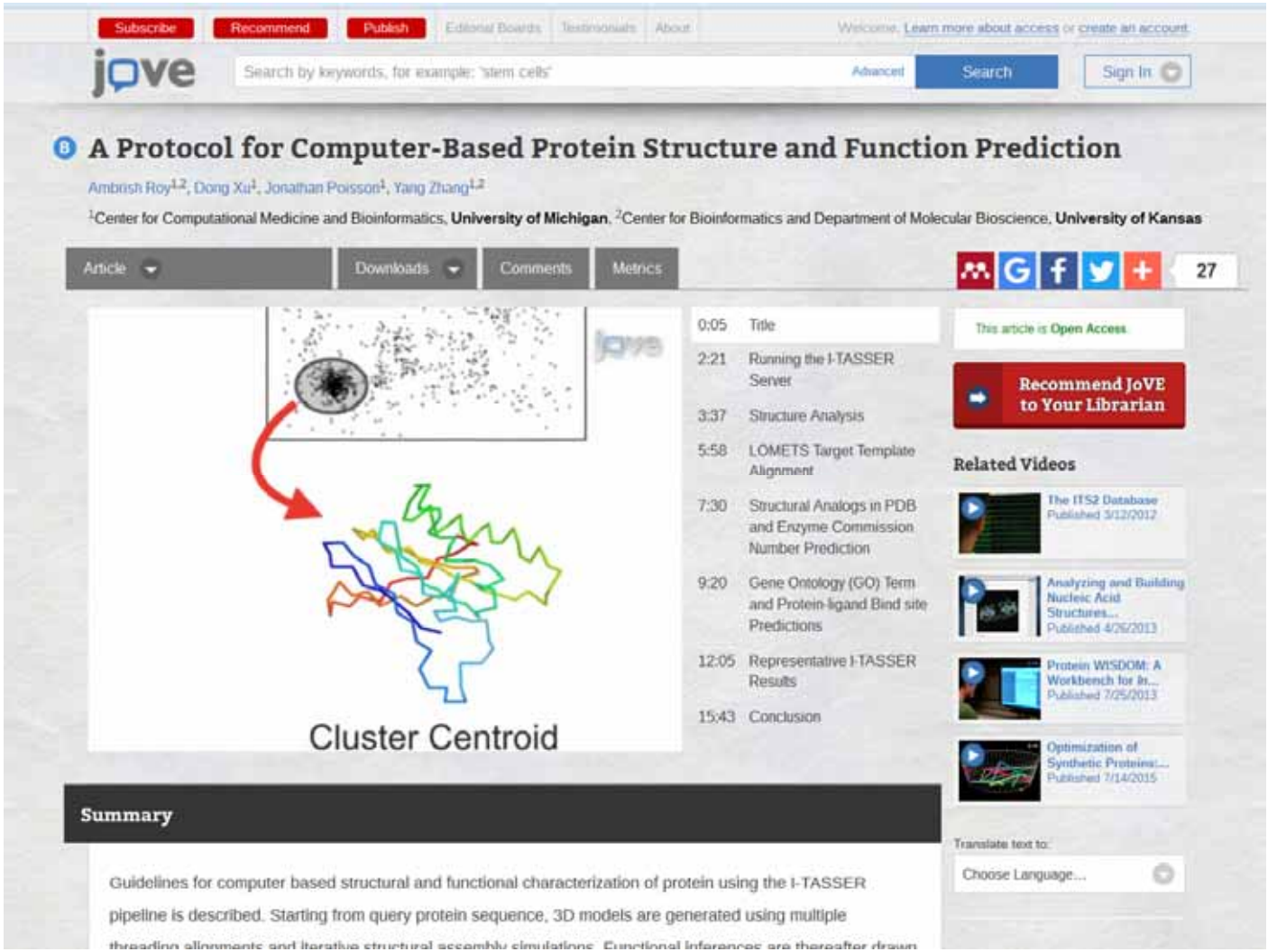
- Discover unexplored interactions in PPI-networks and gene regulatory networks
- Learn the structure
- Reconstruct the structure





Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. & Müller, T. 2008. Identifying functional modules in protein–protein interaction networks: an integrated exact approach. Bioinformatics, 24, (13), i223-i231.

From structure to function



http://www.jove.com/video/3259/a-protocol-for-computer-based-protein-structure-function

Interesting: Hubs tend to link to small degree nodes

Nodes: proteins

Links: physical interactions (binding)

Puzzling pattern:

Hubs tend to link to small degree nodes.

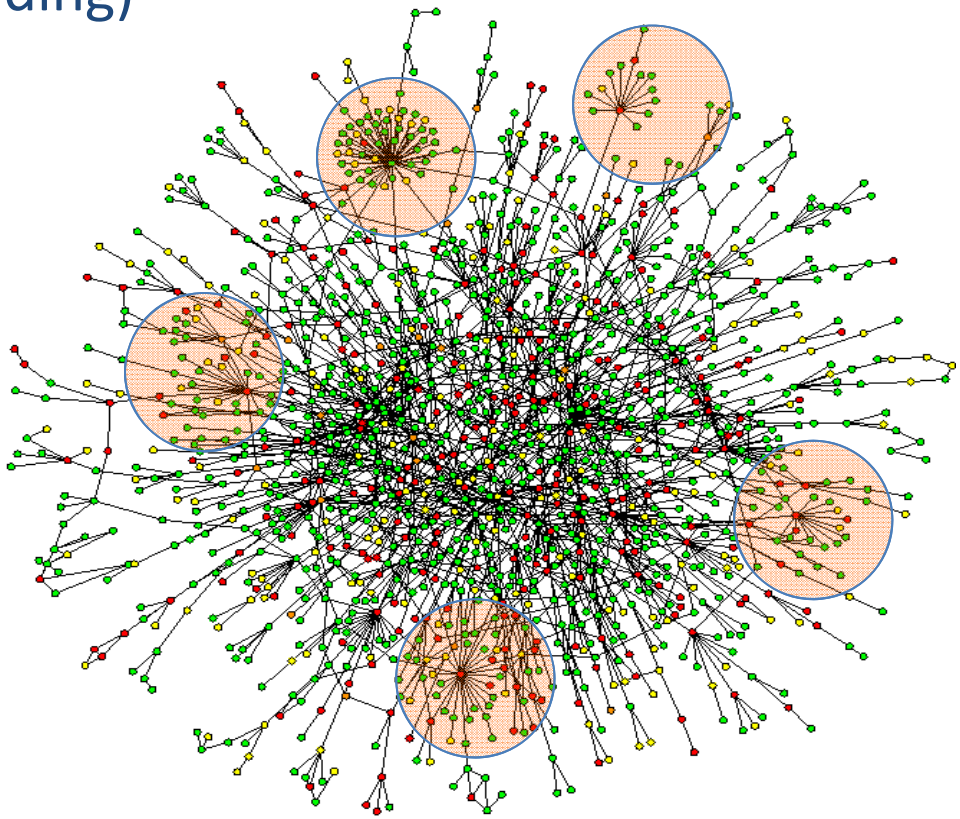
Why is this puzzling?

In a random network, the probability that a node with degree *k* links to a node with degree *k'* is:

$$p_{kk'} = \frac{kk'}{2L}$$

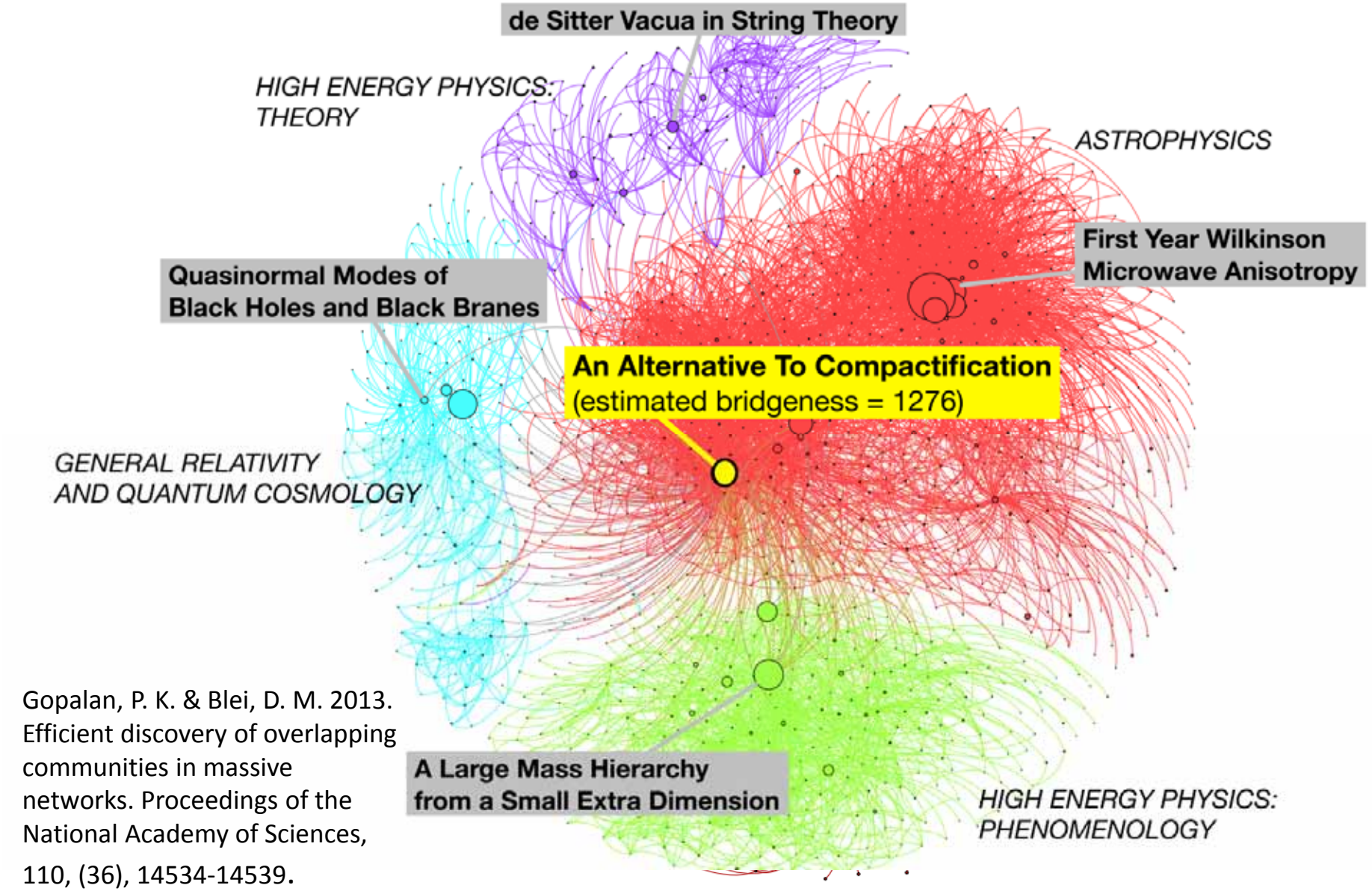
k≅50, k'=13, N=1,458, L=1746

$p_{50,13} = 0.15$ $p_{2,1} = 0.0004$



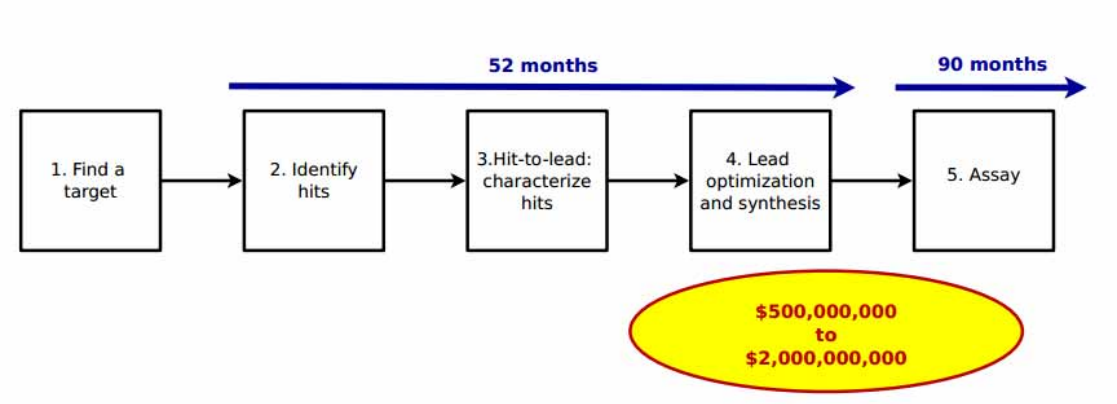
Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. 2001. Lethality and centrality in protein networks. Nature, 411, (6833), 41-42.

Example: Subgraph Discovery



Gopalan, P. K. & Blei, D. M. 2013. Efficient discovery of overlapping communities in massive networks. Proceedings of the National Academy of Sciences, 110, (36), 14534-14539.

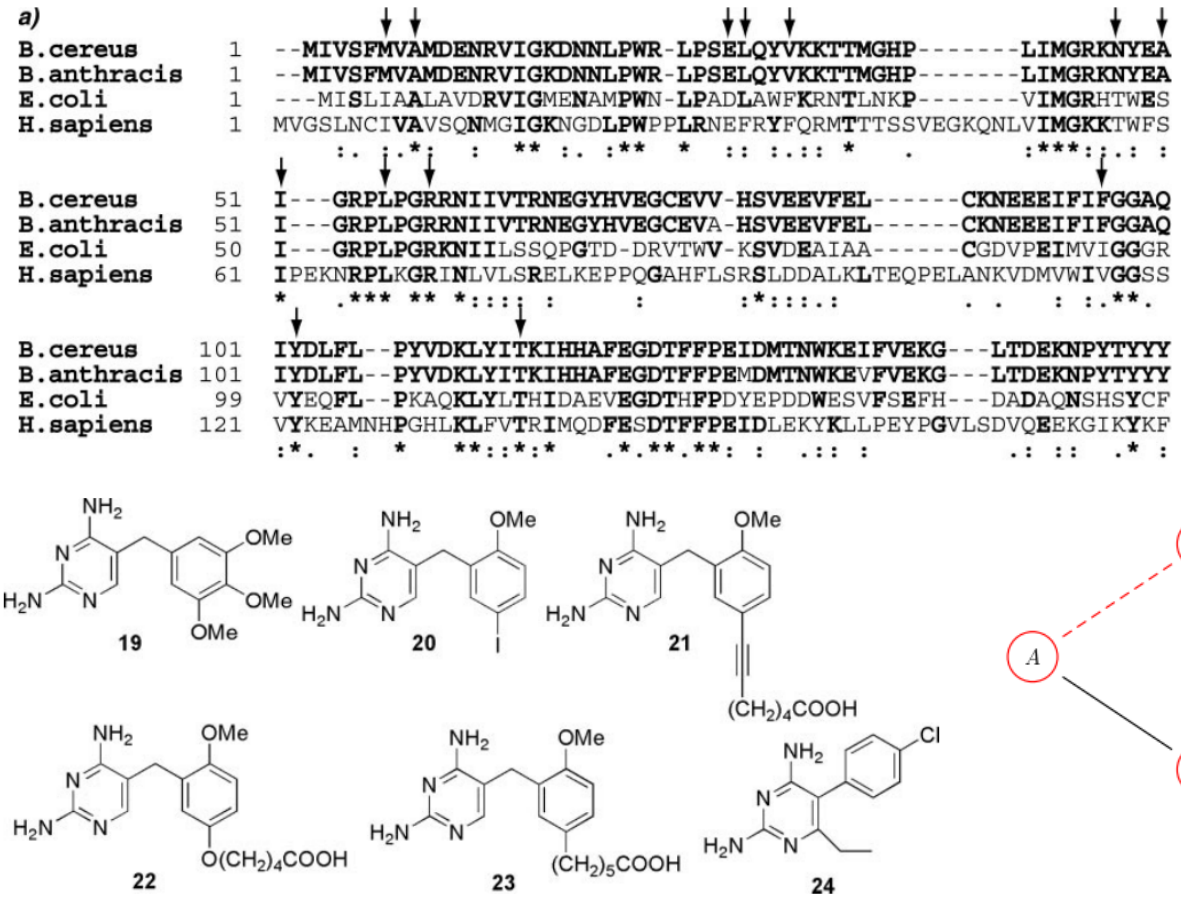
- A) Discovery of unexplored interactions
- B) Learning and Predicting the structure
- C) Reconstructing the structure
- Which joint probability distributions does a graphical model represent?
- How can we learn the parameters and structure of a graphical model?

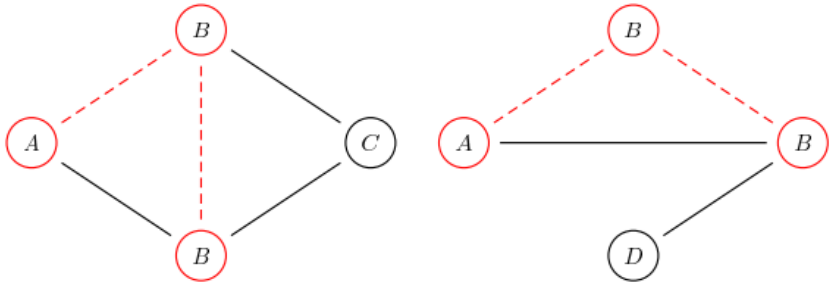


The chemical space

- 10^{60} possible small organic molecules
- 10^{22} stars in the observable universe

Example Question: Predicting Function from Structure





How similar are two graphs? How similar is their structure? How similar are their node and edge labels?

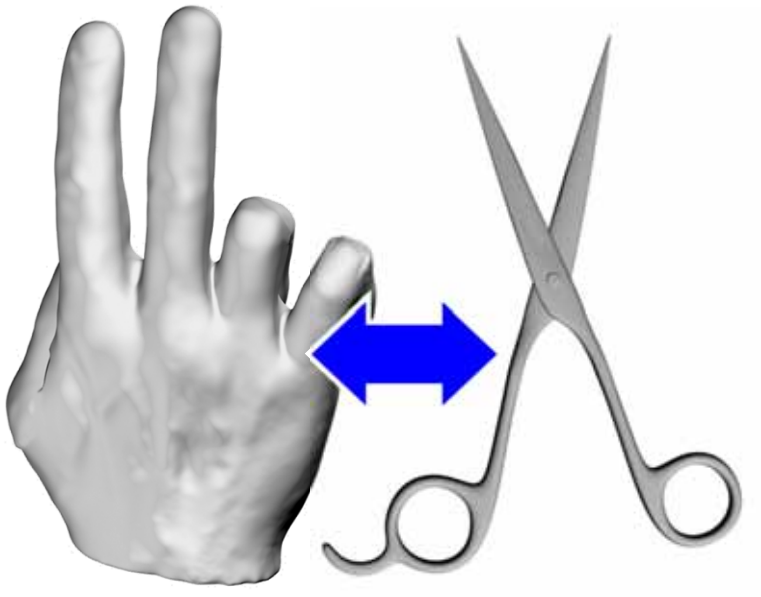
Joska, T. M. & Anderson, A. C. 2006. Structure-activity relationships of Bacillus cereus and Bacillus anthracis dihydrofolate reductase: toward the identification of new potent drug leads. Antimicrobial agents and chemotherapy, 50, 3435-3443.

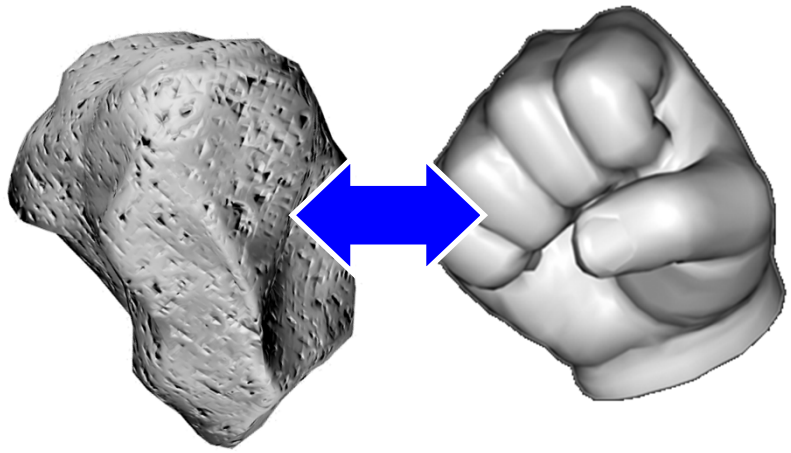
- Similar Property Principle: Molecules having similar structures should have similar activities.
- Structure-based representations: Compare molecules by comparing substructures, e.g.
  - Sets as vectors: Measure similarity by the cosine distance
  - Sets as sets: Measure similarity by the Jaccard distance
  - Sets as points: Measure similarity by Euclidean distance
- Problems: Dimensionality, Non-Euclidean cases

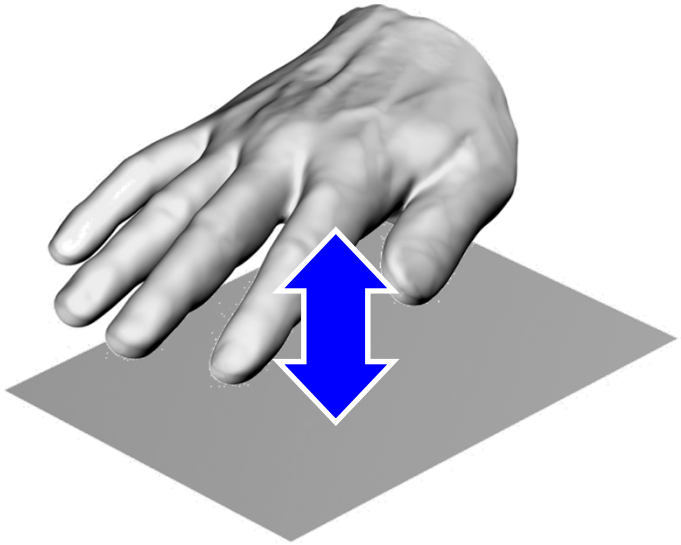
4) Little Excursus: What is similarity?

What is Similar?

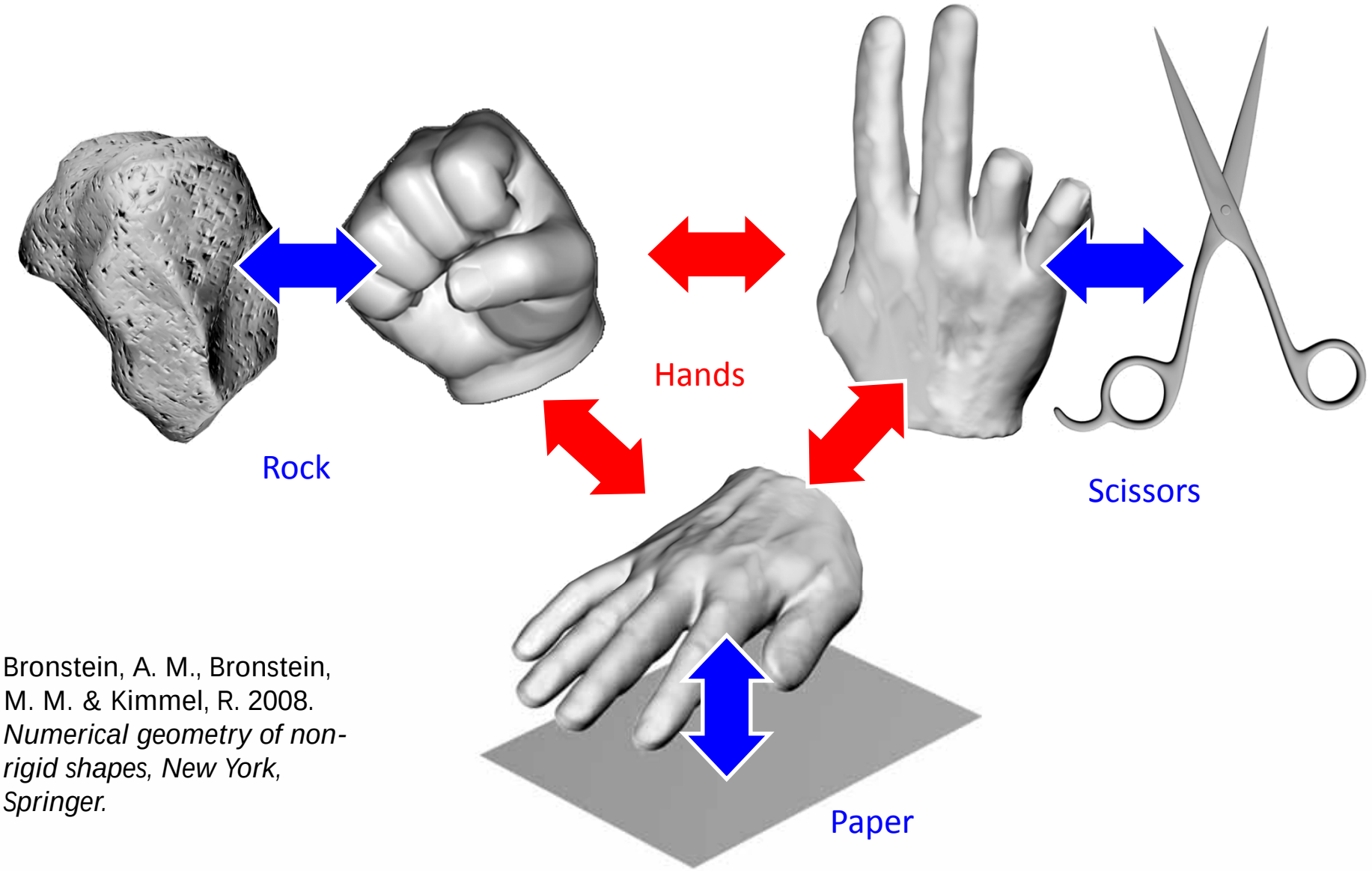
Image credit to Eamonn Keogh (2008)







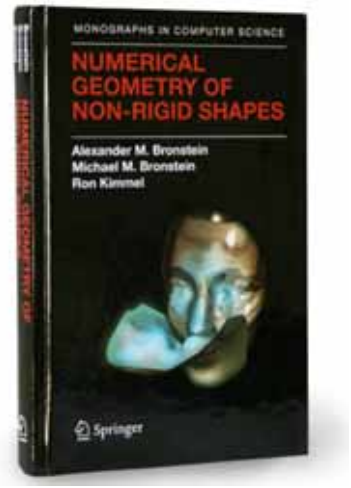
Bronstein, A. M., Bronstein, M. M. & Kimmel, R. 2008. *Numerical geometry of non-rigid shapes, New York, Springer.*



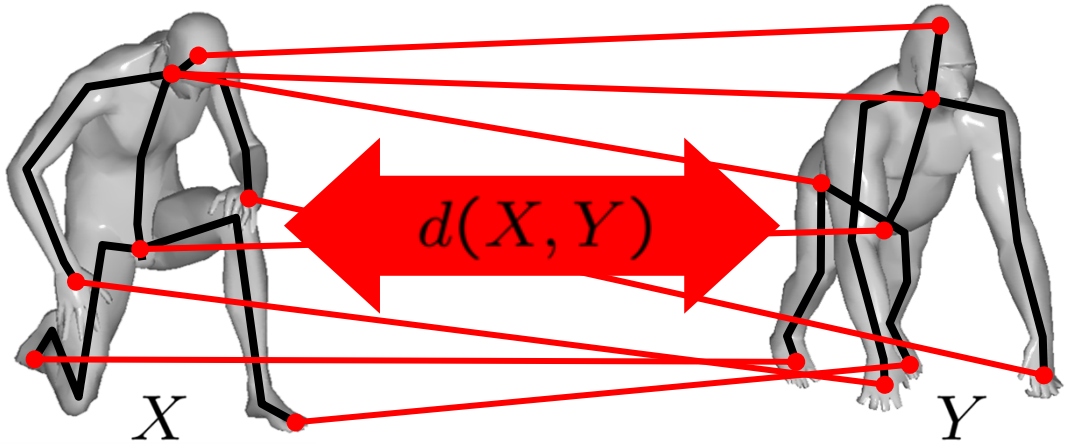
Bronstein, A. M., Bronstein, M. M. & Kimmel, R. 2008. *Numerical geometry of non-rigid shapes, New York, Springer.*

Similarity and Correspondence

Bronstein, A. M., Bronstein, M. M. & Kimmel, R. 2008. *Numerical geometry of non-rigid shapes, New York, Springer.*



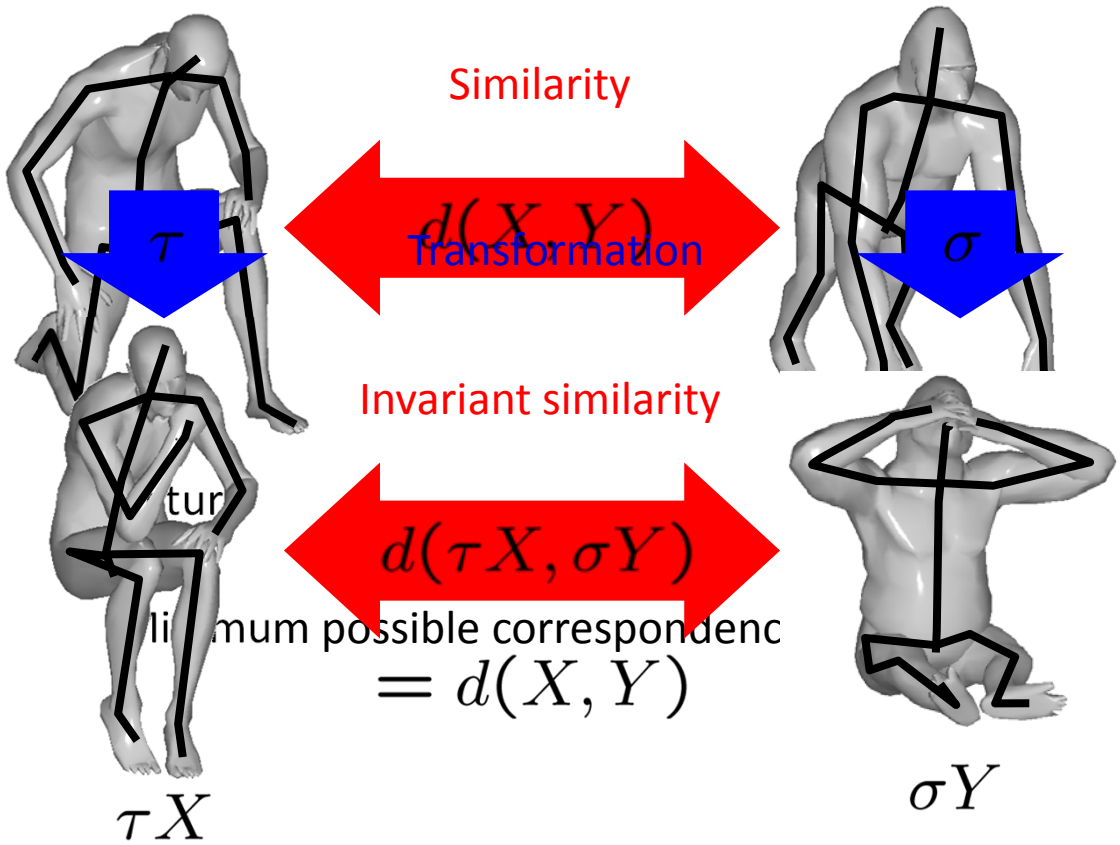
http://www.inf.usi.ch/bronstein/



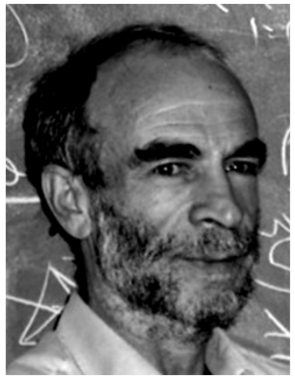
Structure Structure

Correspondence quality = structure similarity (distortion)

Minimum possible correspondence distortion

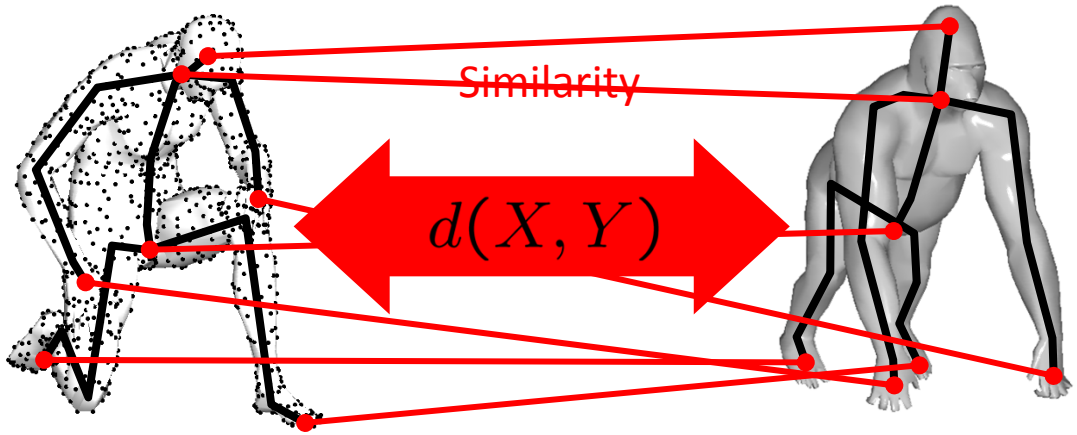

Similarity
$d(X,Y)$
Transformation
τ
σ
Invariant similarity
tur
$d(\tau X, \sigma Y)$
li mum possible correspondenc
$= d(X,Y)$
τX
σY

Gromov-Hausdorff dist: finding the opt. correspondence



Michail Gromov
(1943-)

Gromov, M. (1984) Infinite groups as geometric objects.



Metric space

Metric space

Felix Hausdorff
(1868-1942)

$$d_{\mathrm{GH}}(X,Y) = \frac{1}{2} \min_{\mathcal{C}} \max_{\substack{(x_i,y_i)\in\mathcal{C} \\ (x_j,y_j)\in\mathcal{C}}} |\delta_X(x_i,x_j) - \delta_Y(y_i,y_j)|$$

$$\forall x_i \, \exists y_i \;\; \text{s.t.} (x_i, y_i) \in \mathcal{C} \qquad \forall y_i \, \exists x_i \;\; \text{s.t.} (x_i, y_i) \in \mathcal{C}$$

Discrete optimization over correspondences is NP hard !

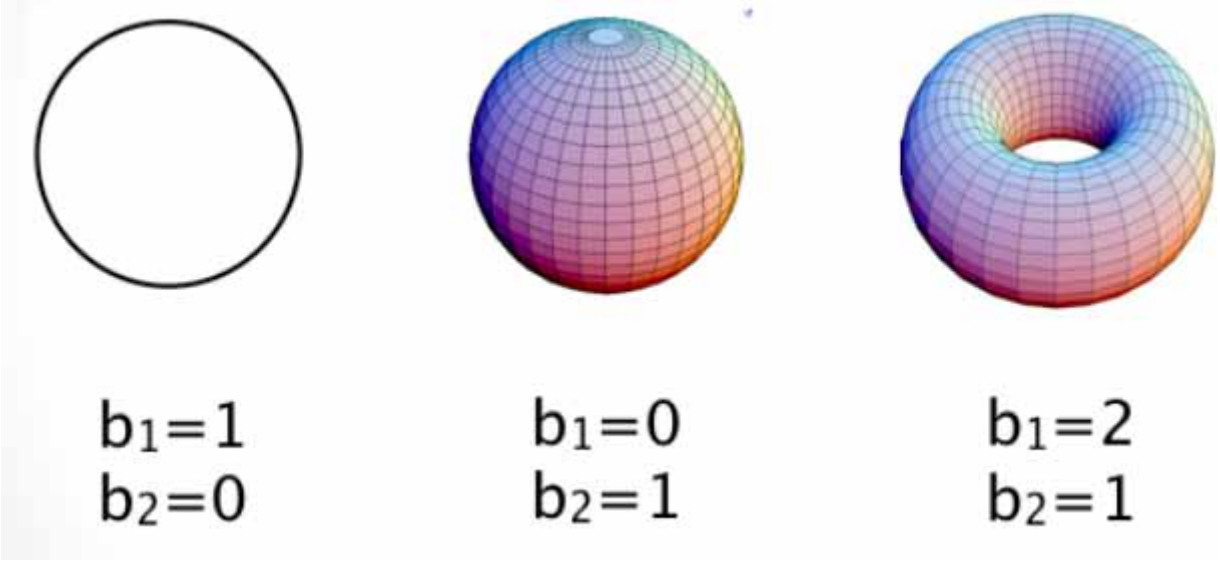
Distinguish topological spaces

Enrico Betti
(1823-1892)

Counts the number of “i-dimensional holes”

bi is the “i-th Betti number”

Emmy Noether
(1882-1935)



Betti numbers are computed as dimensions of Boolean vector spaces (E. Noether)

Zomorodian, A. & Carlsson, G. 2005. Computing Persistent Homology. *Discrete & Computational Geometry, 33, (2), 249-274.*

Structural Patterns are often hidden in weakly str. data

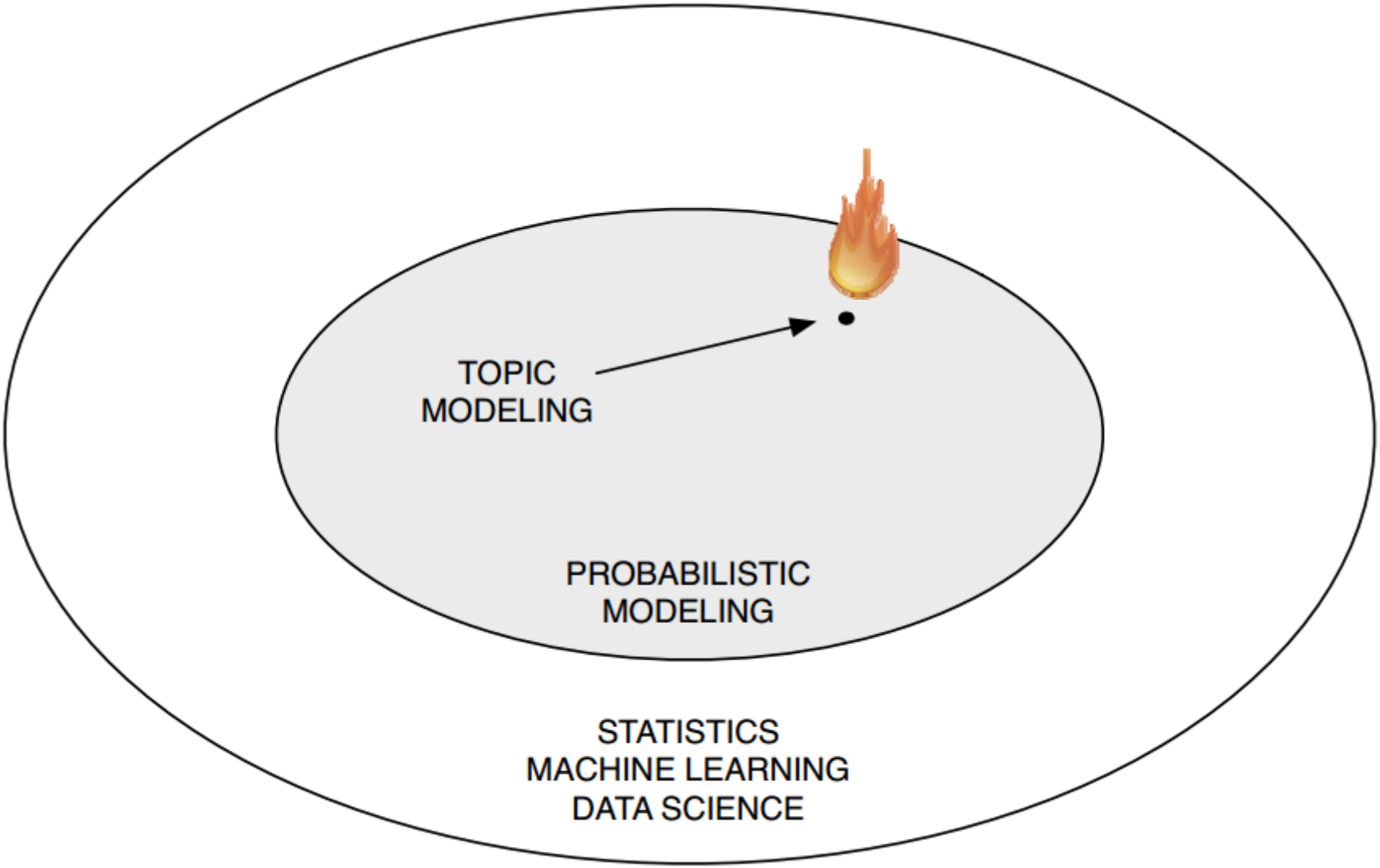
- Statement of Vin de Silva (2003), Pomona College:
- Let M be a topological or metric space, known as the *hidden parameter space*;
- let $\mathbb{R}^d$ be a Euclidean space, the *observation space*,
- and let $f: M \longrightarrow \mathbb{R}^d$ be a continuous embedding.
- Furthermore, let $X \subset M$ be a finite set of data points, perhaps the realization of a stochastic process, i.e., a family of random variables $\{X_i, i \in I\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, and denote $Y = f(X) \subset \mathbb{R}^d$ the images of these points under the mapping f.
- We refer to X as *hidden data*, and Y as the *observed data*.
- M, f and X are unknown, but Y is - so can we identify M?

- Mega Problem: To date none of our known methods, algorithms and tools scale to the massive amount and dimensionalities of data we are confronted in practice;
- we need much more research efforts towards making computational topology successful as a general method for data mining and knowledge discovery

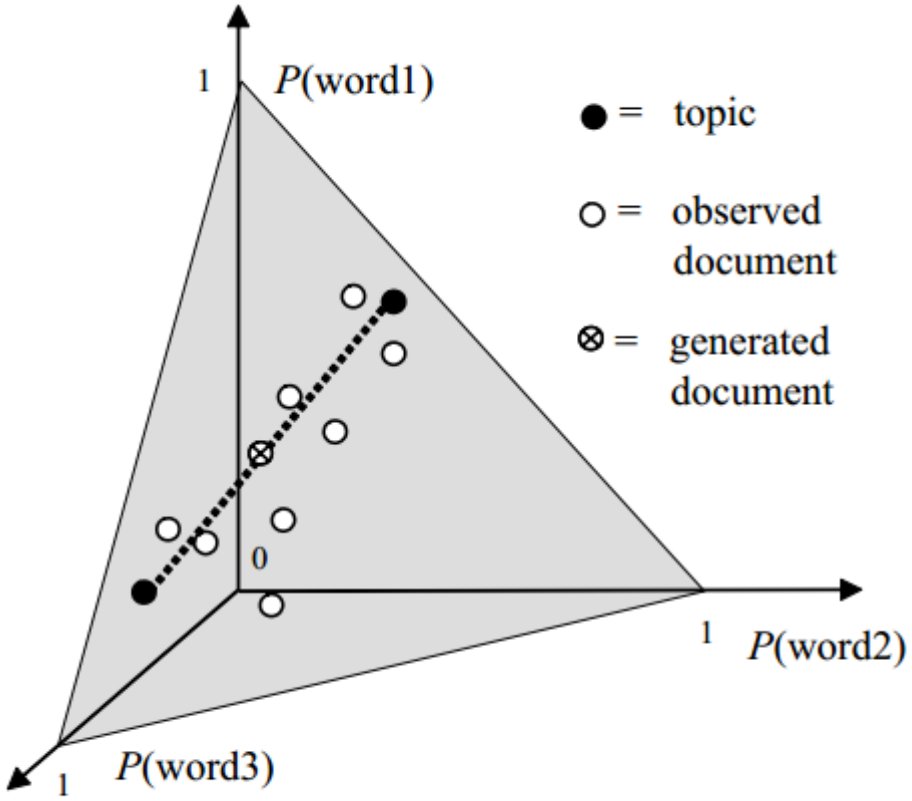
Holzinger, A. 2014. On Topological Data Mining. In: Lecture Notes in Computer Science, LNCS 8401. Berlin Heidelberg: Springer, pp. 331-356, doi:10.1007/978-3-662-43968-5_19.

5) Probabilistic Topic Models

Topic modelling – small topic but hot topic in ML



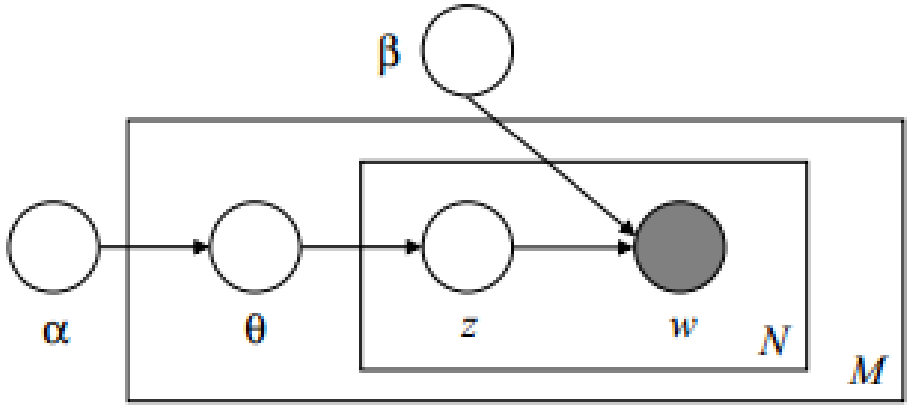
Geometry of Topic Models

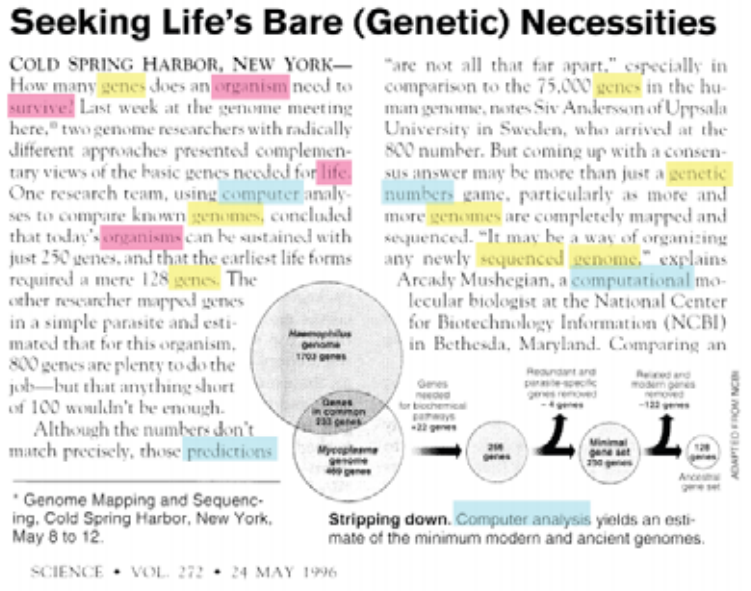


- Documents are categorical distributions over some predefined vocabulary of (10,000+) words
- Topics are categorical distributions on same vocabulary
- Generative model: Each document is (nearly) a convex combination of the topic distributions

D. Blei, 2008

Generative statistical model for natural language





Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions "are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

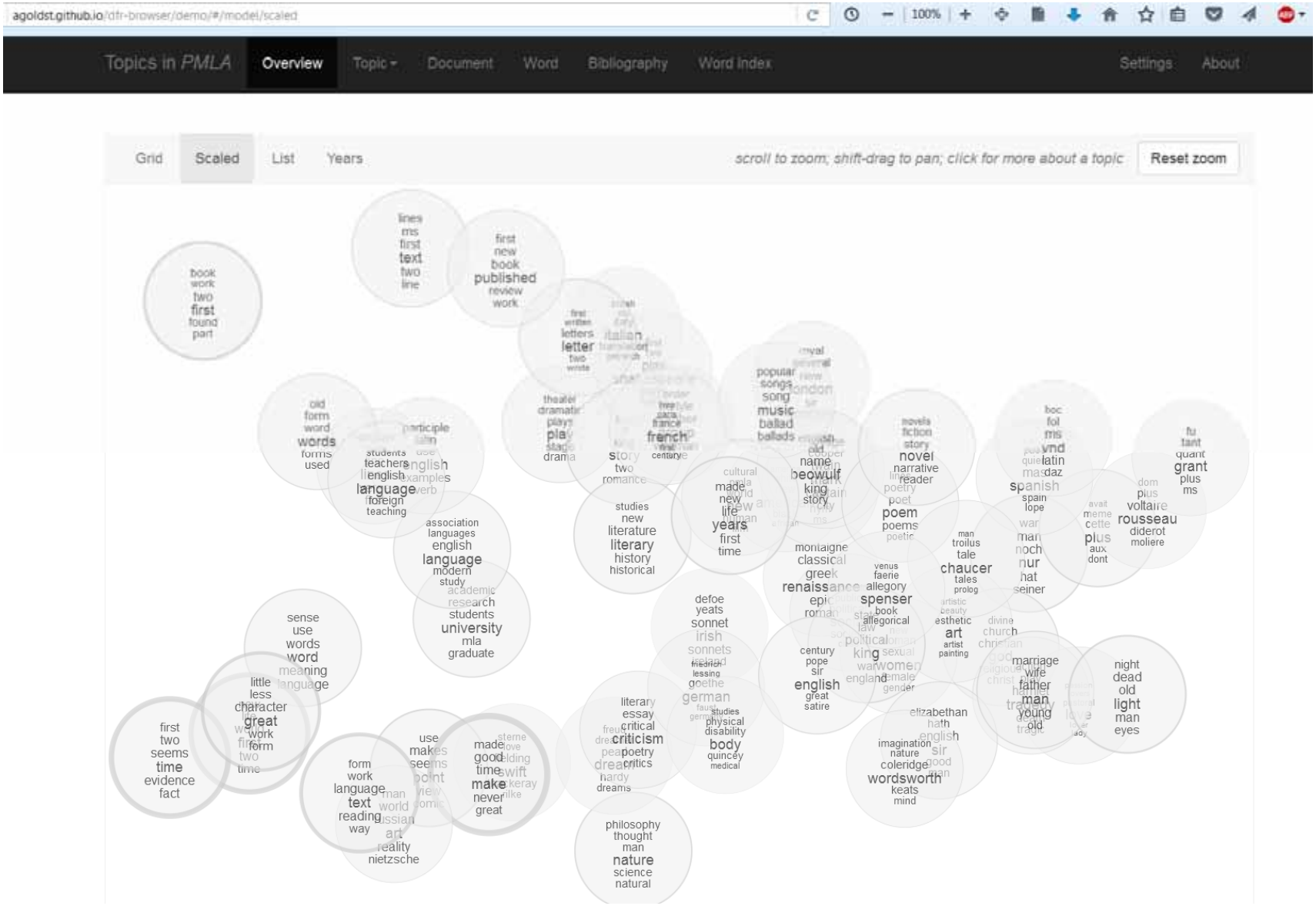
SCIENCE • VOL. 272 • 24 MAY 1996

Given the parameters α and β, the joint distribution of a topic mixture θ, a set of N topics z, and a set of N words w is given by:

$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta)$$

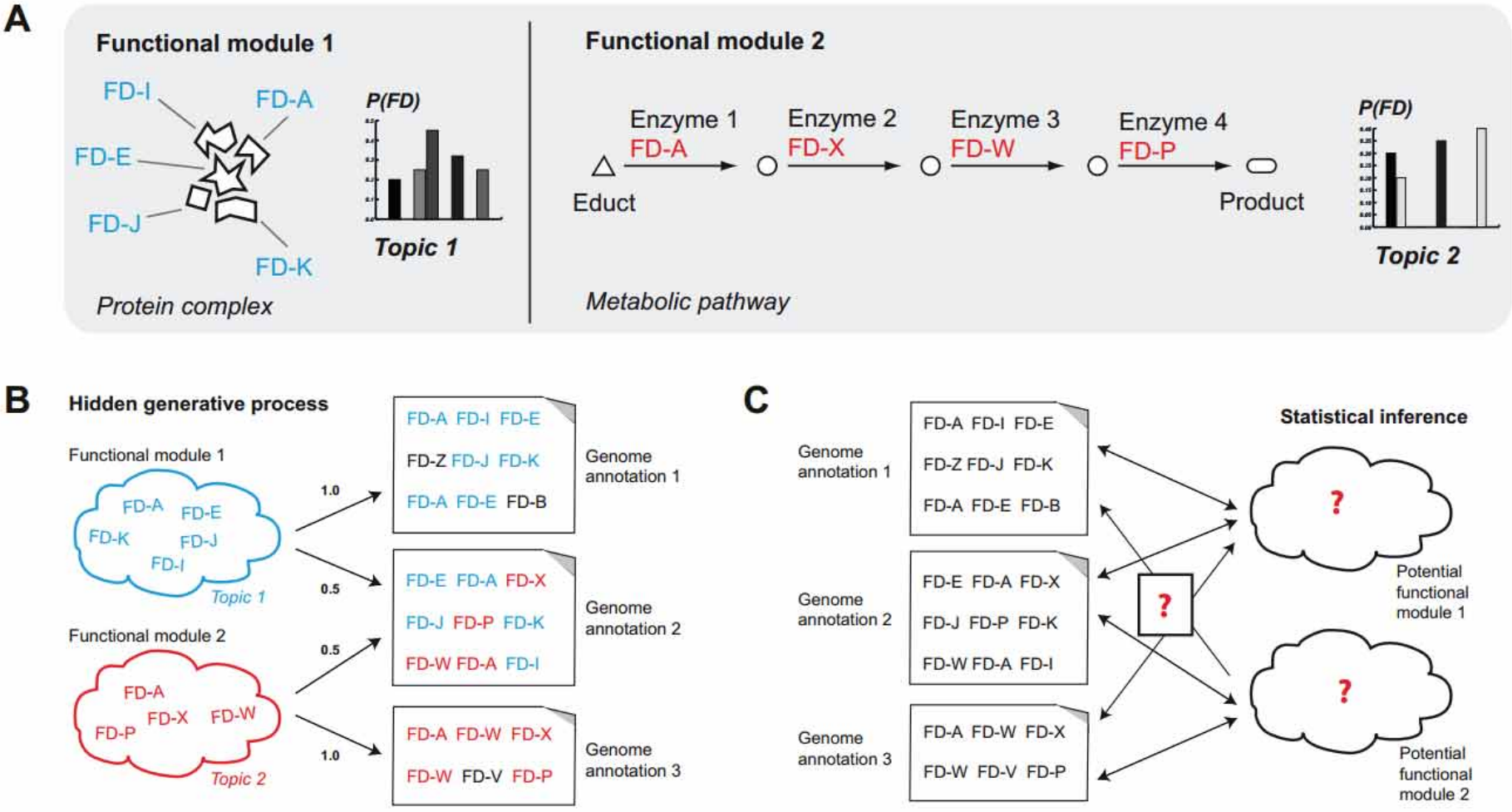
Blei, D. M., Ng, A. Y. & Jordan, M. I. 2003. Latent dirichlet allocation. The Journal of machine Learning research, 3, 993-1022.

Motivation: to get insight into unknown document sets



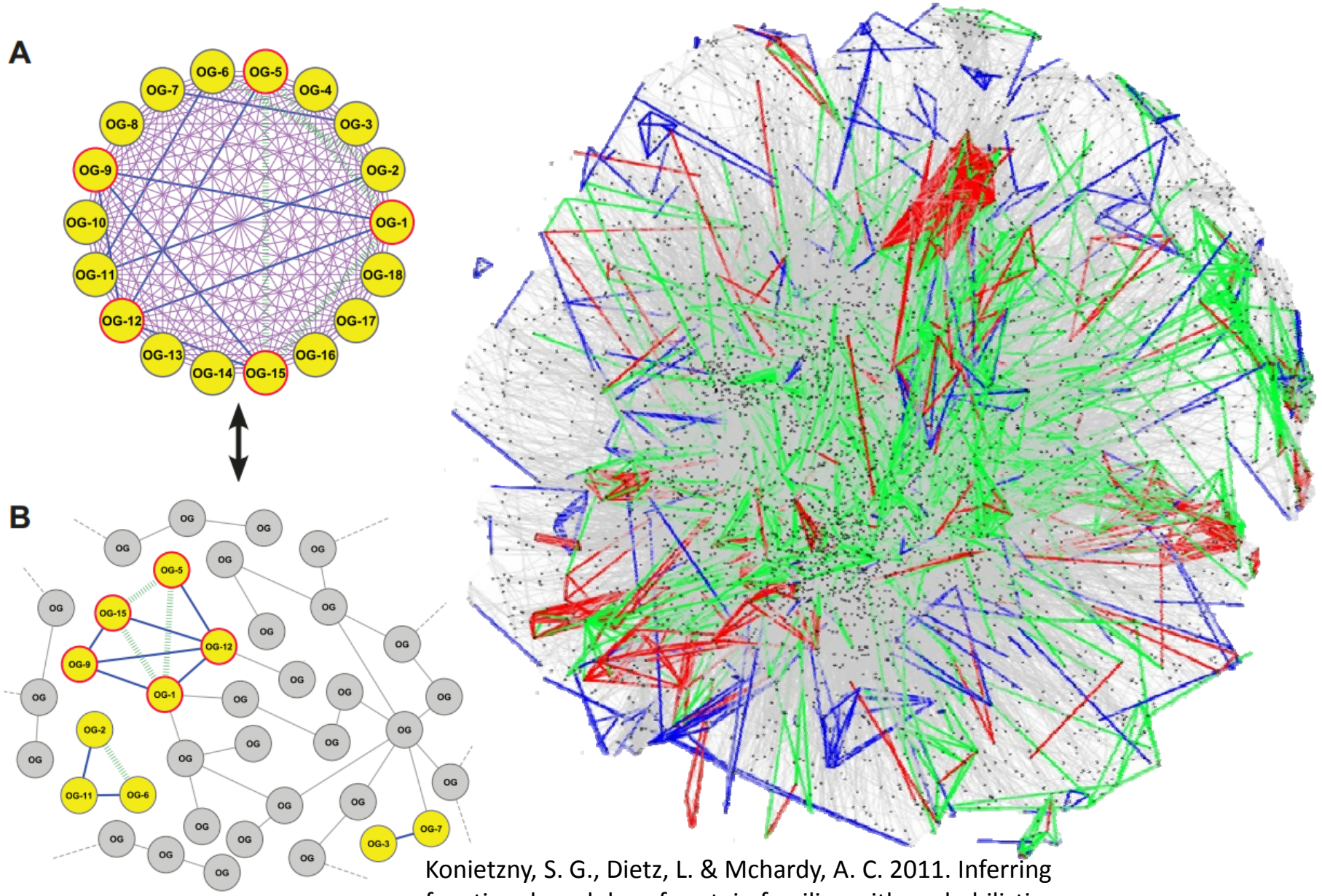
http://agoldst.github.io/dfr-browser/demo/#/model/scaled

Example from Bioinformatics



Konietzny, S. G., Dietz, L. & Mchardy, A. C. 2011. Inferring functional modules of protein families with probabilistic topic models. BMC bioinformatics, 12, (1), 1.

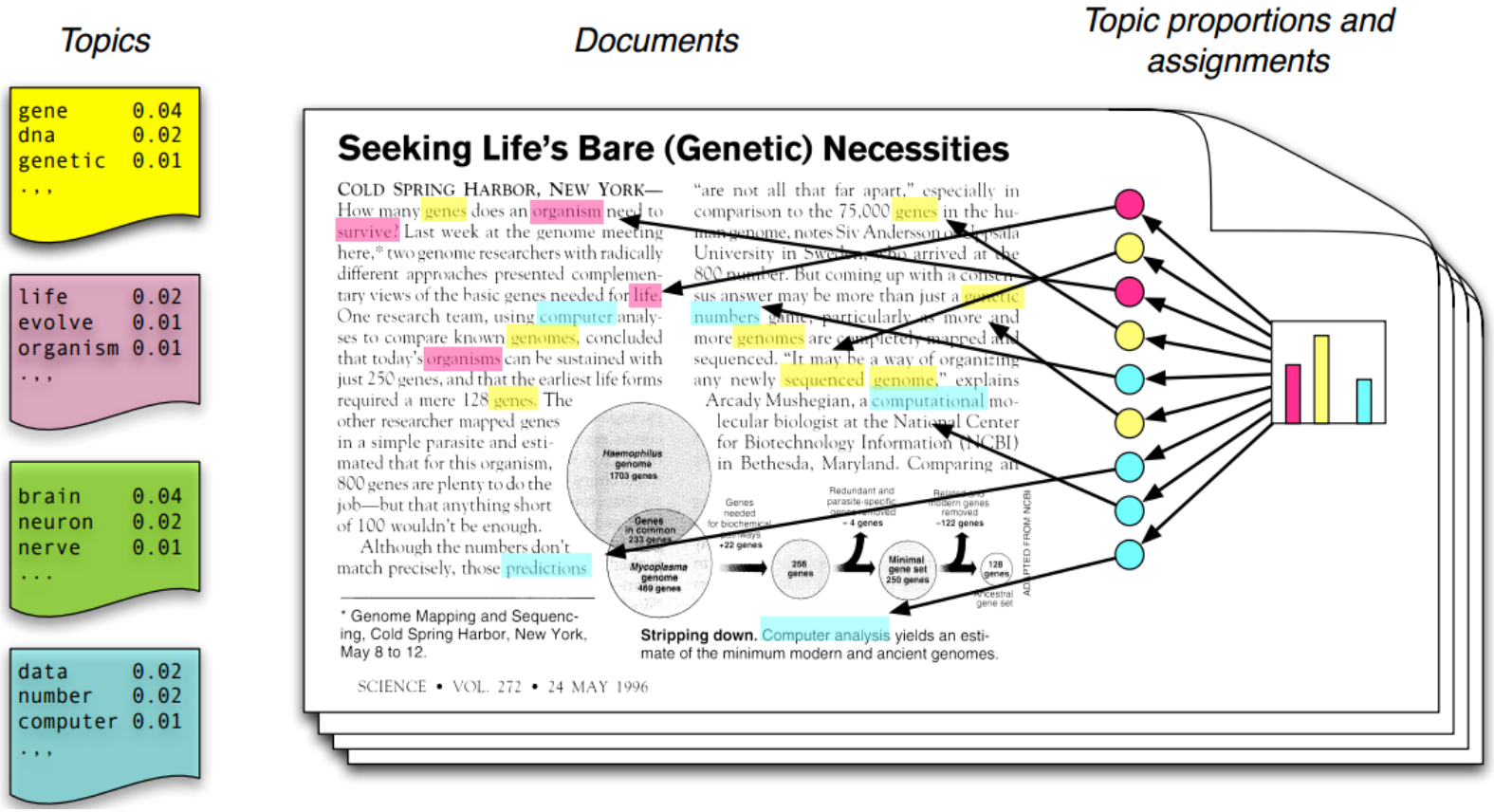
Eval. scheme for inferred potential functional modules



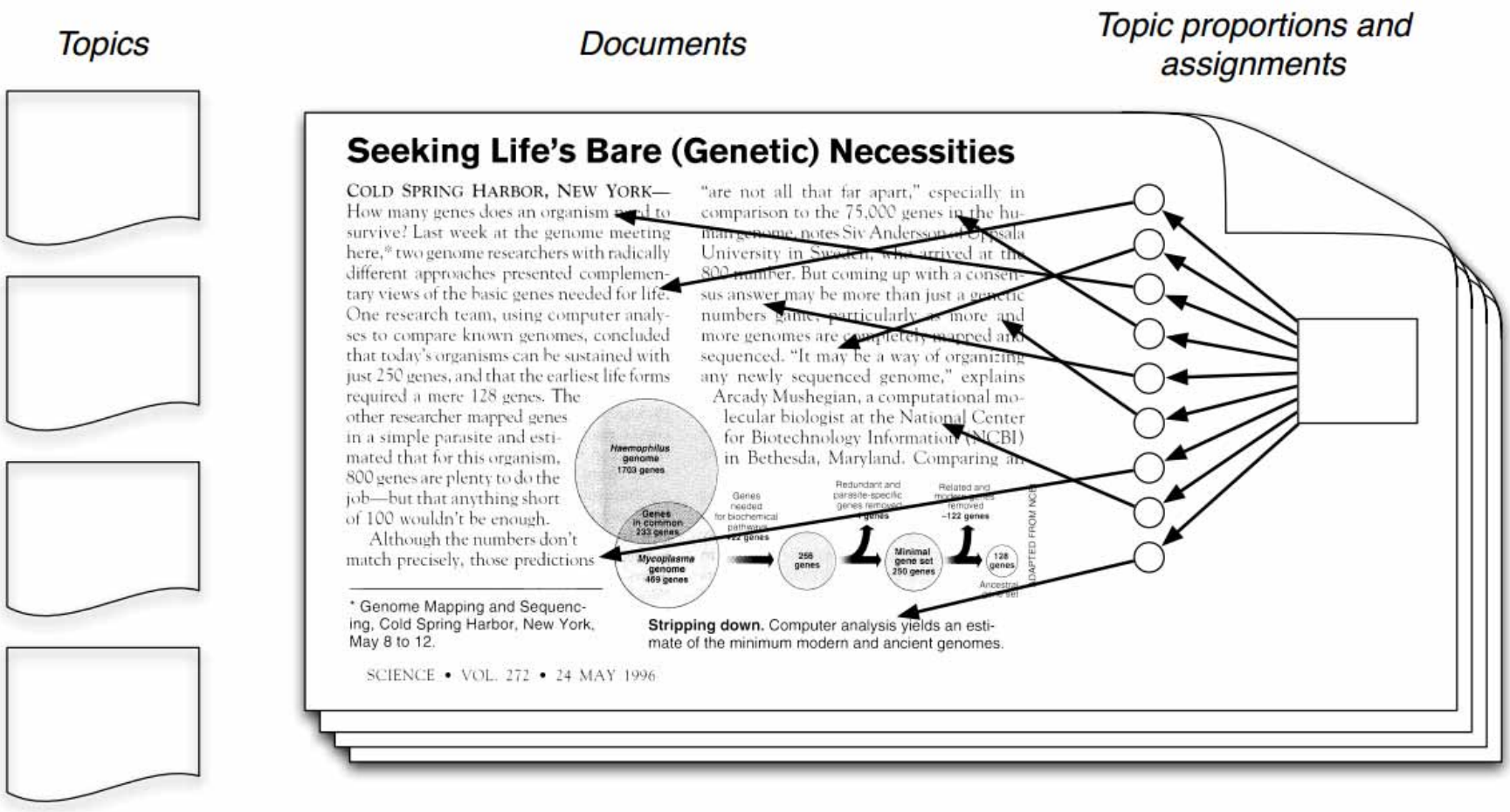
Konietzny, S. G., Dietz, L. & Mchardy, A. C. 2011. Inferring functional modules of protein families with probabilistic topic models. BMC bioinformatics, 12, (1), 1.

Goal: to get insight in unknown document collections

See a nice demo http://agoldst.github.io/dfr-browser/demo/#/model/grid



Each doc is a random mix of corpus-wide topics and each word is drawn from one of these topics

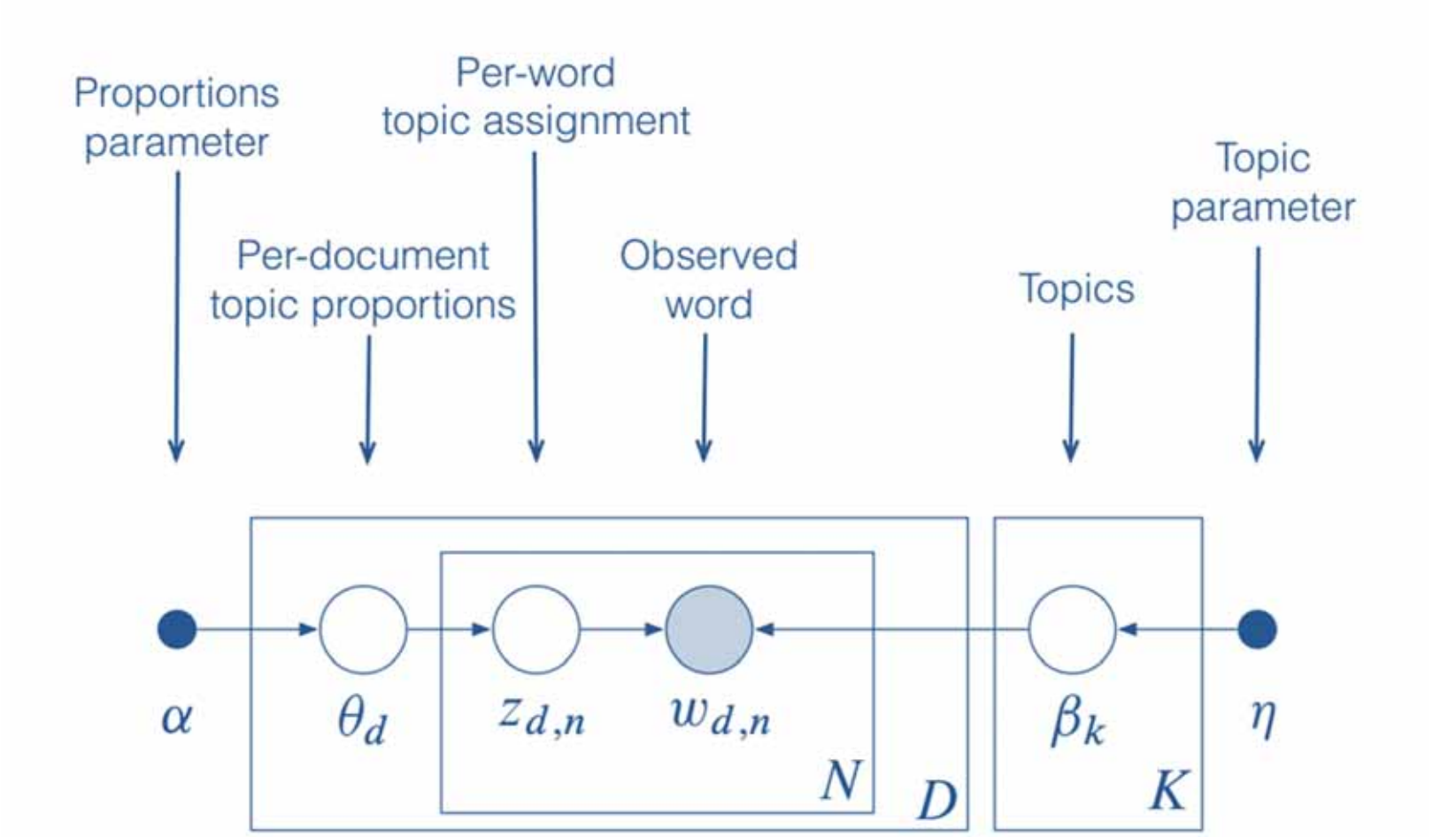


We only observe the docs – the other structure is hidden; then we compute the posterior p(t,p,a|docs)

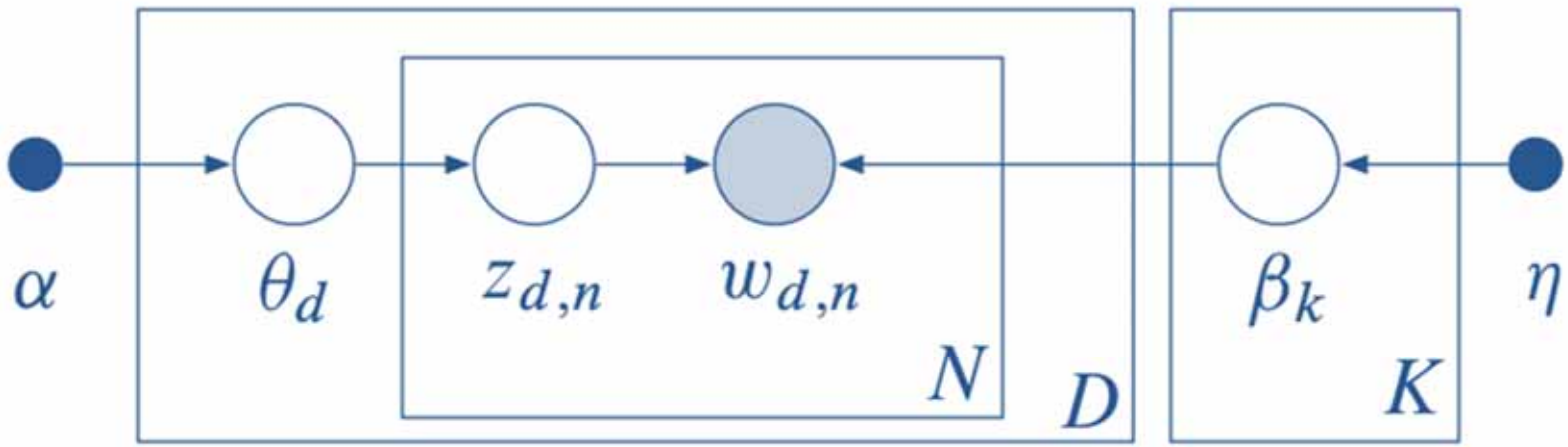
Output Example: 4 learned topics

human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

LDA is an example for a probabilistic graphical model



- Encodes assumptions on data with a factorization of the joint
- Connects assumptions to algorithms for computing with data
- Defines the posterior (through the joint)

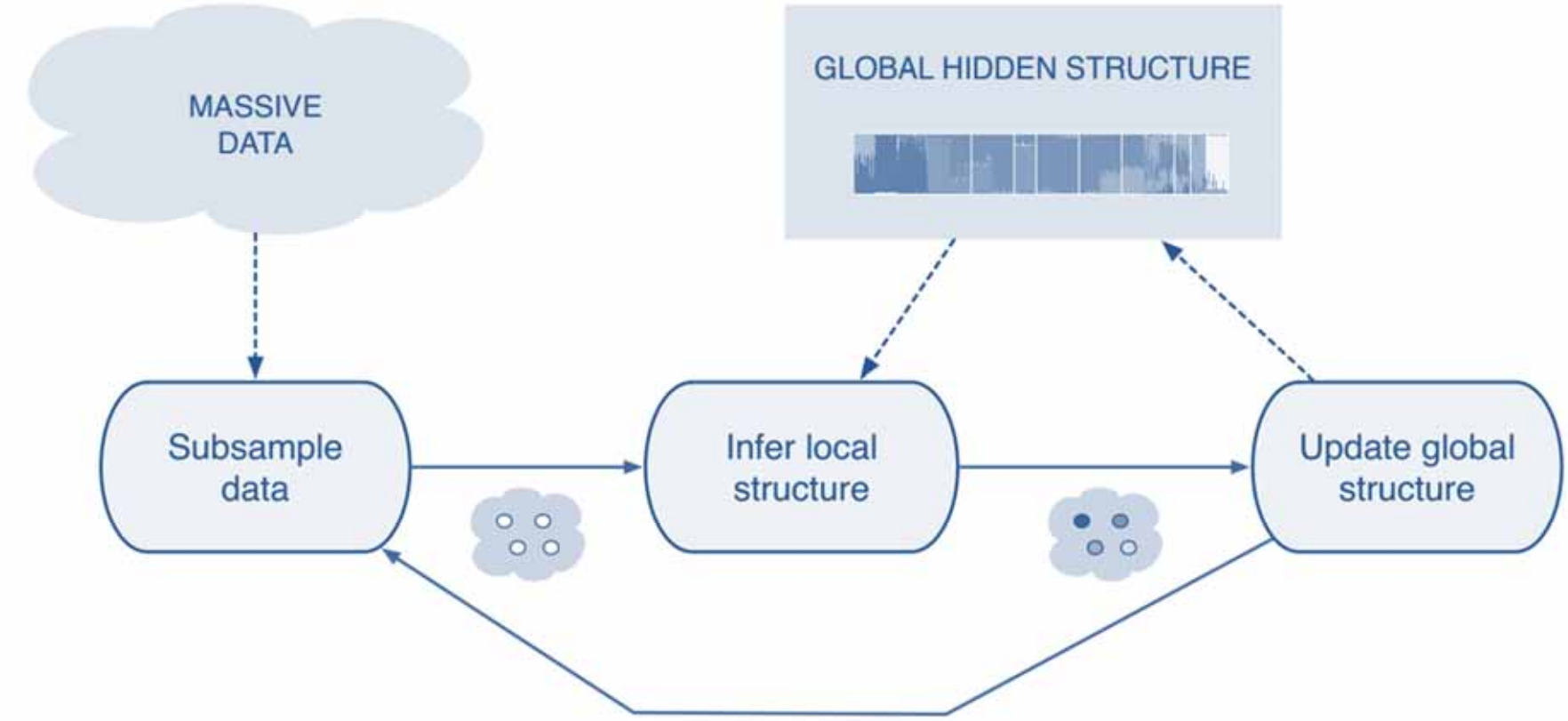


$$p(\beta, \boldsymbol{\theta}, \mathbf{z} \mid \mathbf{w}) = \frac{p(\beta, \boldsymbol{\theta}, \mathbf{z}, \mathbf{w})}{\int_{\beta} \int_{\boldsymbol{\theta}} \sum_{\mathbf{z}} p(\beta, \boldsymbol{\theta}, \mathbf{z}, \mathbf{w})}$$

We can't compute the denominator, the marginal $p\ (w)$, therefore we use approximate inference;

However, this do not scale well …

For “big data” stochastic variational inference



Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. 2013. Stochastic variational inference. The Journal of Machine Learning Research, 14, (1), 1303-1347.

1: Initialize $\lambda^{(0)}$ randomly.
2: Set the step-size schedule ρ_t appropriately.
3: **repeat**
4: Sample a document w_d uniformly from the data set.
5: Initialize $\gamma_{dk} = 1$, for $k \in \{1, \ldots, K\}$.
6: **repeat**
7: For $n \in \{1, \ldots, N\}$ set

$$\phi_{dn}^k \propto \exp\left\{\mathbb{E}[\log \theta_{dk}] + \mathbb{E}[\log \beta_{k,w_{dn}}]\right\}, \; k \in \{1, \ldots, K\}.$$

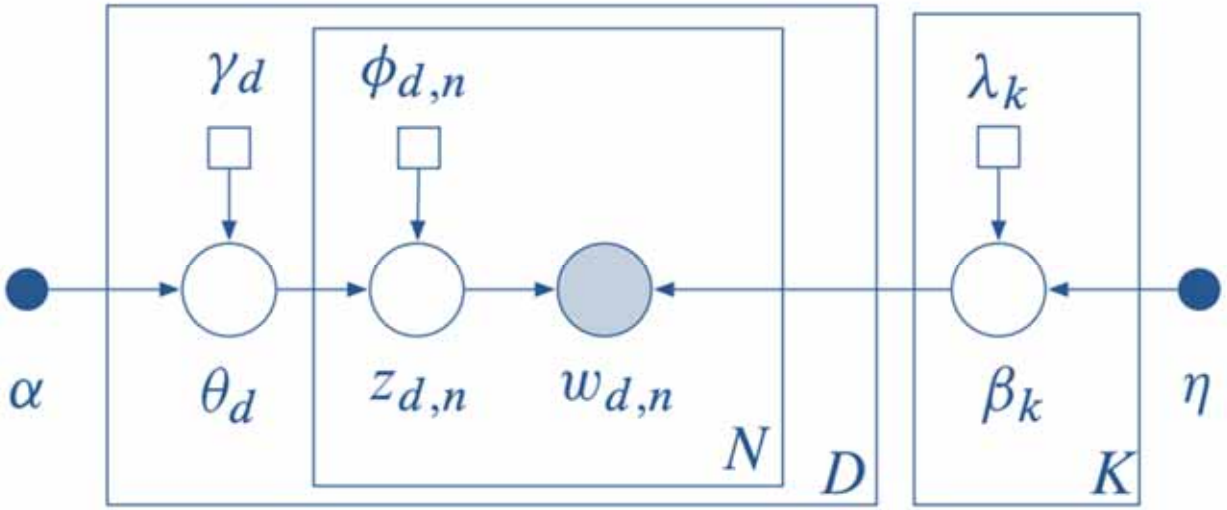
8: Set $\gamma_d = \alpha + \sum_n \phi_{dn}$.
9: **until** local parameters ϕ_{dn} and γ_d converge.
10: For $k \in \{1, \ldots, K\}$ set intermediate topics

$$\hat{\lambda}_k = \eta + D \sum_{n=1}^{N} \phi_{dn}^k w_{dn}.$$

11: Set $\lambda^{(t)} = (1 - \rho_t)\lambda^{(t-1)} + \rho_t \hat{\lambda}$.
12: **until** forever

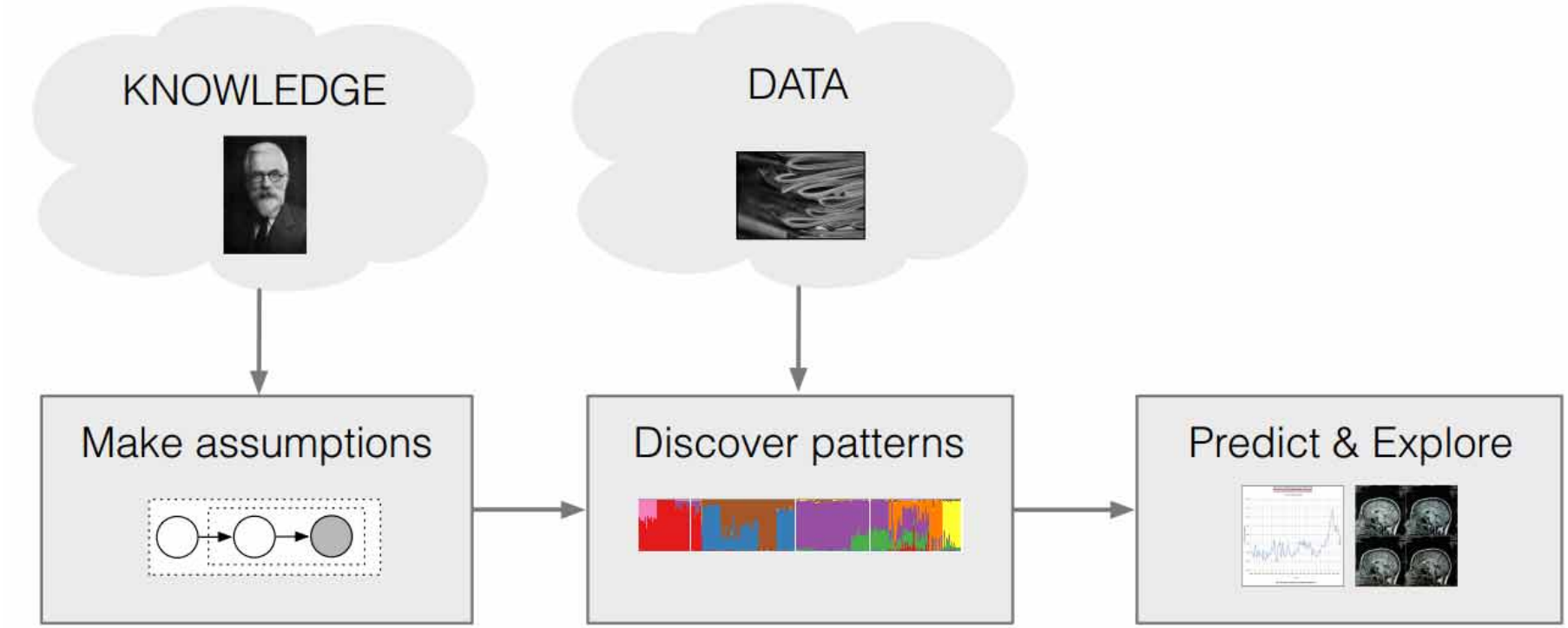
Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. 2013. Stochastic variational inference. The Journal of Machine Learning Research, 14, (1), 1303-1347.

Stochastic variational inference in LDA



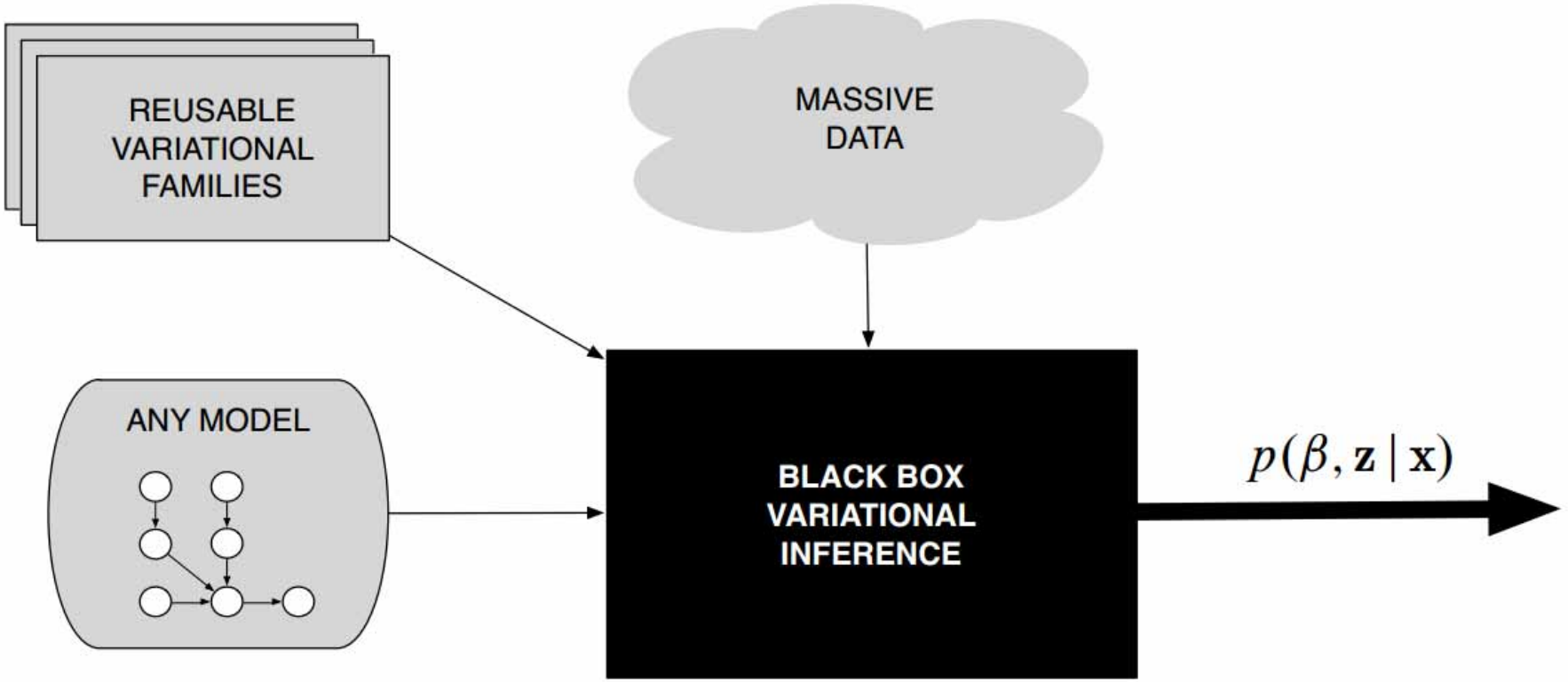
1. Sample a document
2. Estimate the local variational parameters using the current topics
3. Form intermediate topics from those local parameters
4. Update topics as a weighted average of intermediate and current topics

Approximate inference can be difficult to achieve



- Approximate inference can be difficult to derive.
- Especially true for models that are not conditionally conjugate (Discrete choice models, Bayesian generalized linear models, ...)
- Holds us back from trying many models.

Black Box Approach



- Easily use variational inference with *any model*
- No exponential family requirements
- No mathematical work beyond specifying the model

Conclusion: What is needed …

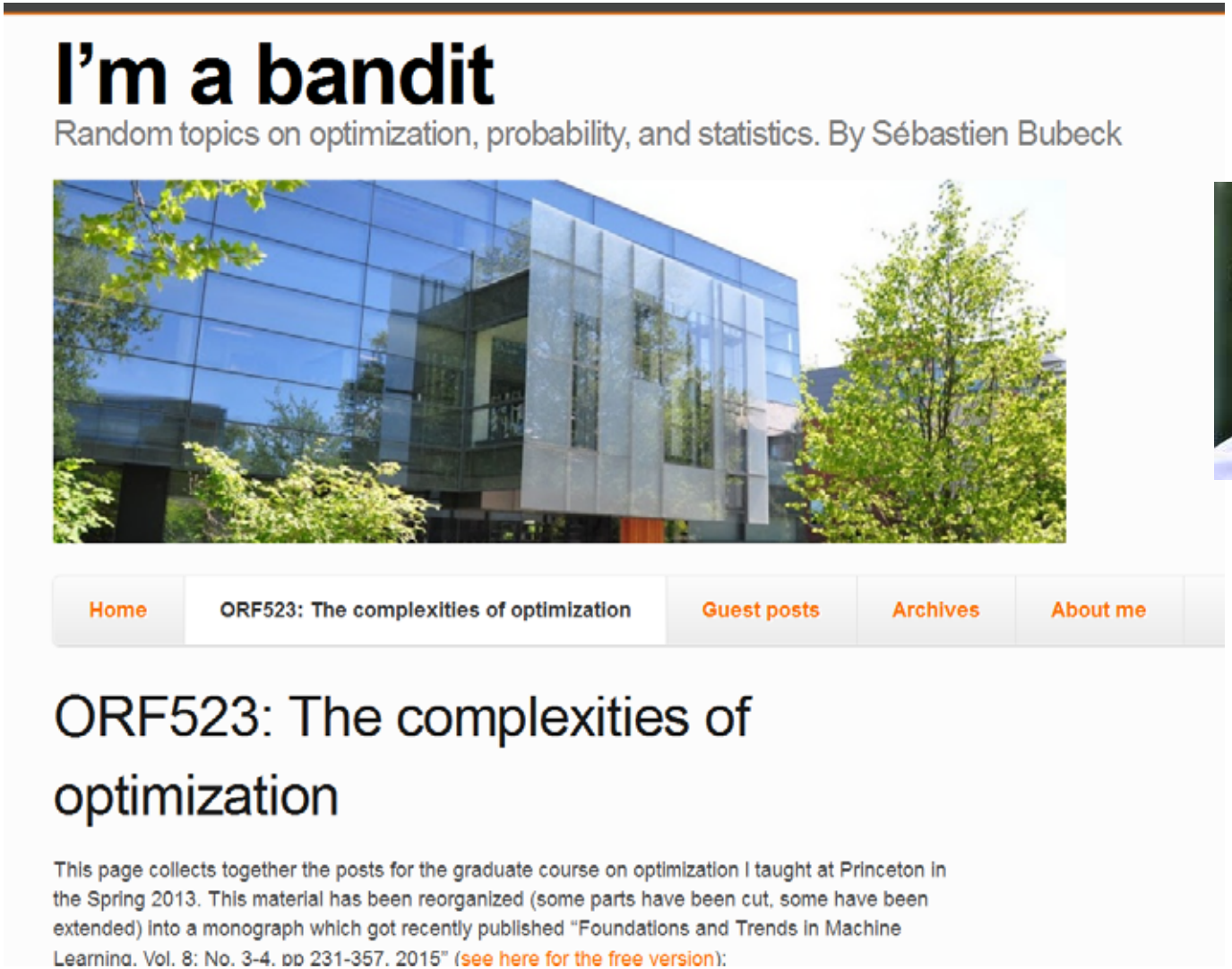
- Flexible and expressive components for building models
- Scalable and generic inference algorithms
- Easy to use software to stretch probabilistic modeling into new areas
- Topic models are one approach towards detection of topics in text collections
- More general: Identify re-occurring patterns in data collections

Topic model toolkits

- Particular topic models
  - Stanford topic model toolbox
    http://nlp.stanford.edu/software/tmt
  - Topic modeling at Princeton
    http://www.cs.princeton.edu/~blei/topicmodeling.html
  - MALLET (Java) http://mallet.cs.umass.edu
  - Network topic models: Bayes-stack
    https://github.com/bgamari/bayes-stack
  - Gensim (Python) http://radimrehurek.com/gensim/
  - R package for Topic models. http://epub.wu.ac.at/3987/
- Frameworks for generative models
  - Variational inference: Infer.net
    http://research.microsoft.com/infernet/
  - Gibbs sampling: OpenBUGS http://openbugs.net/

6) Graph Bandits

The complexities of optimization: Sébasitien Bubeck



Bubeck, S. & Cesa-Bianchi, N. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. Machine Learning, 5, (1), 1-122.

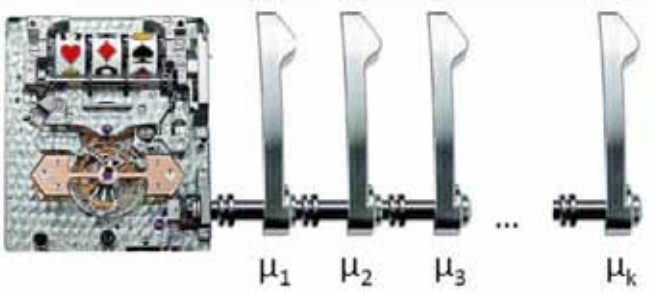
https://blogs.princeton.edu/imabandit/

Also very interesting: Bubeck, S. 2015. Convex optimization: Algorithms and complexity. Foundations and Trends in Machine Learning, 8, (3-4), 231-357.

Remember Bandits

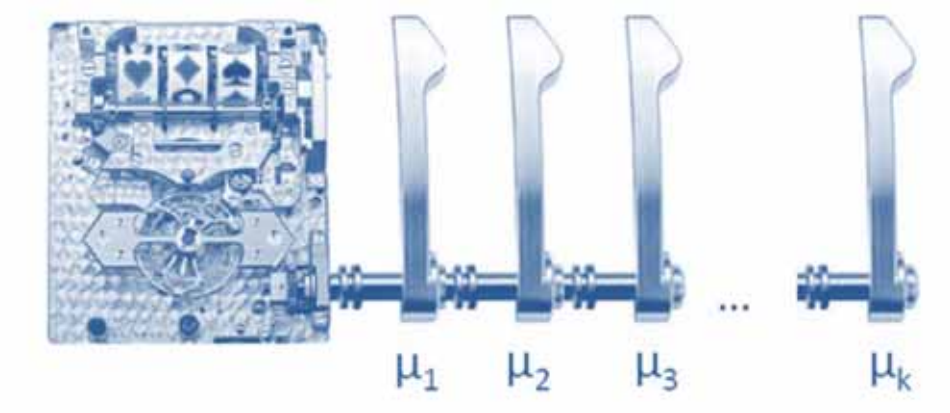
- Slot-machine (bandit - robs your money)
- One-armed bandit
- Very simple model for sequential decision making under uncertainty
- Main challenge: exploration versus exploitation
- Many application domains: A/B-Testing, Crowdsourcing, optimization, search, …

Multi-Armed Bandits problem



- Multi-armed bandit:= a gambler strategically operating multiple machines in order to draw the highest possible profits
- There are n slot-machines ("einarmige Banditen")
- Each machine i returns a reward $y \approx P(y; \Theta_i)$
- Challenge: The machine parameter Θ_i is unknown
- Which arm of a slot machine should a gambler pull to maximize his cumulative reward over a sequence of trials? (stochastic setting or adversarial setting)

Machine Parameters of the k-armed Bandit



Each arm a either

wins (reward=1) with fixed (unknown) probability μ_a, or

loses (reward=0) with fixed (unknown) probability $1 - \mu_a$

- All draws are independent given μ_1 ... μ_k
- Problem:
  How to pull arms to maximize the total reward?

Underlying Principle of the k-Armed Bandits problem

- Let $a_t \in \{1, \dots, n\}$ be the choice of a machine at time t
- Let $y_t \in \mathbb{R}$ be the outcome with a mean of $\langle y_{at} \rangle$
- Now, the given policy maps all history to a new choice:

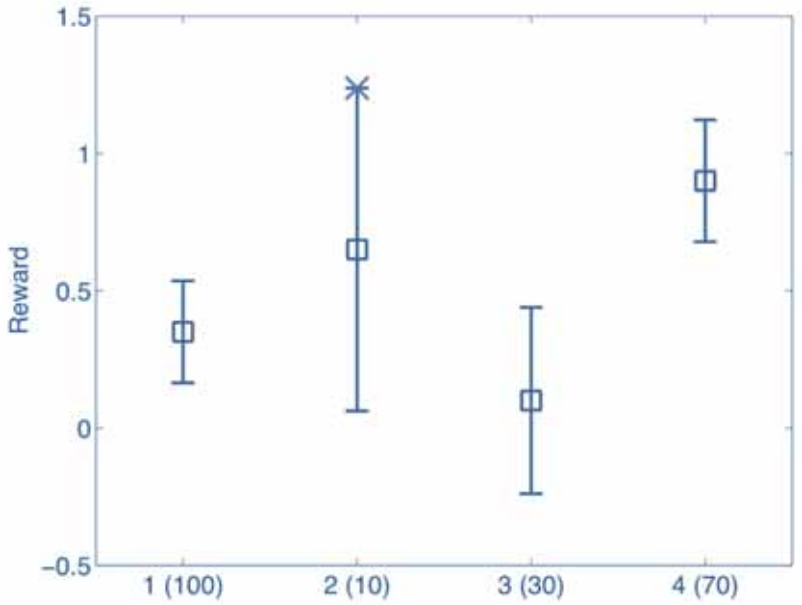
$$\pi : [(a_1, y_1), (a_2, y_2), \dots, (a_{t-1}, y_{t-1})] \mapsto a_t$$

- The problem: Find a policy π that $\max\langle y_T \rangle$
- Now, two effects appear when choosing such machine:
  - You collect more data about the machine (=knowledge)
  - You collect reward
- Exploration and Exploitation
  - **Exploration:** Choose the next action a_t to $min\langle H(b_t) \rangle$
  - **Exploitation:** Choose the next action a_t to $max\langle y_t \rangle$
- models an agent that simultaneously attempts to acquire new knowledge (called "exploration") and optimize his or her decisions based on existing knowledge (called "exploitation"). The agent attempts to balance these competing tasks in order to maximize total value over the period of time considered.

More information: http://research.microsoft.com/en-us/projects/bandits

MAP-Principle: "Optimism in the face of uncertainty"

$$a_t = \max_{a \in \mathcal{A}} \left(\widehat{r}_t(a) + \sqrt{\frac{\log(1/\delta)}{T_t(a)}} \right)$$



$$a_t = \max_{a \in \mathcal{A}} \left(\text{rew}_t(a) + \text{uncert}_t(a) \right)$$

Exploitation
the higher the (estimated) reward the higher the chance to select the action

Exploration
the higher the (theoretical) uncertainty the higher the chance to select the action

Auer, P., Cesa-Bianchi, N. & Fischer, P. 2002. Finite-time analysis of the multiarmed bandit problem. Machine learning, 47, (2-3), 235-256.

A bandit in a graph is still a bandit ☺

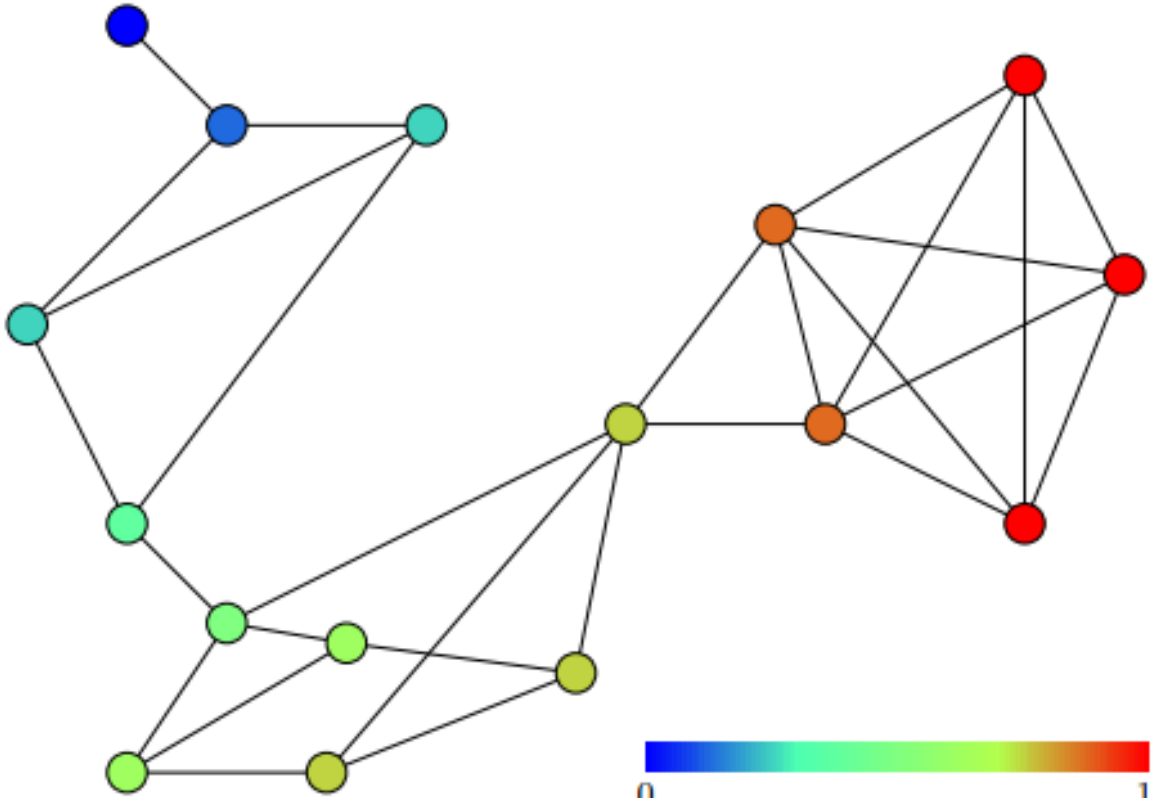
- Let G a known graph with K nodes $\{1, 2, \ldots, K\}$
- Let f be a unknown function defined on the set of nodes
- For $t = 1$ to n,
  - Select a node I_t
  - Observe reward $r_t = f(I_t) + \epsilon_t$
- Goal: maximize sum of expected rewards
- Equivalently minimize regret:

$$R_n = \sum_{t=1}^{n} (f^* - f(I_t)),$$

where $f^* = \max_{1 \leq i \leq K} f(i)$.

- We care about the case when $K > n$

Smooth Graph Function

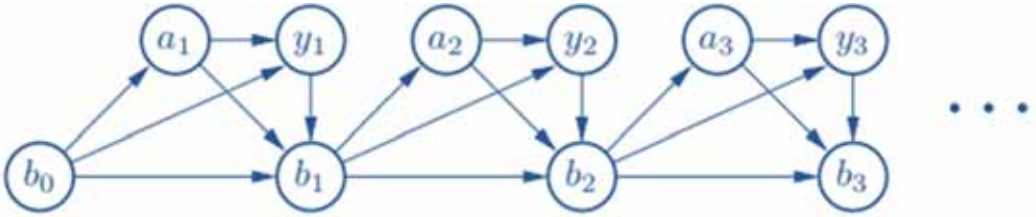


- Knowledge can be represented in two ways:
- 1) as full history $h_t = [(a_1, y_1), (a_2, y_2), \ldots, (a_{t-1}, y_{t-1})]$

  or
- 2) as belief $b_t(\theta) = P(\theta | h_t)$

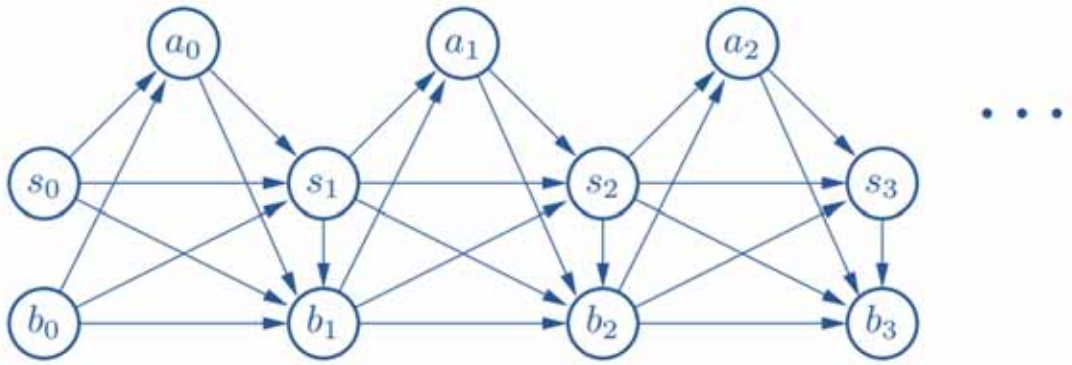
where Θ are the unknown parameters of all machines

The process can be modelled as belief MDP:



$$P(b'|y,a,b) = \begin{cases} 1 & \text{if } b' = b'_{[b,a,y]} \\ 0 & \text{otherwise} \end{cases}, \quad P(y|a,b) = \int_{\theta_a} b(\theta_a)\, P(y|\theta_a)$$

The optimal policies can be modelled as belief MDP



$$P(b'|s',s,a,b) = \begin{cases} 1 & \text{if } b' = b[s',s,a] \\ 0 & \text{otherwise} \end{cases}, \quad P(s'|s,a,b) = \int_\theta b(\theta)\, P(s'|s,a,\theta)$$

$$V(b,s) = \max_a \left[\mathsf{E}(r|s,a,b) + \sum_{s'} P(s'|a,s,b)\, V(s',b') \right]$$

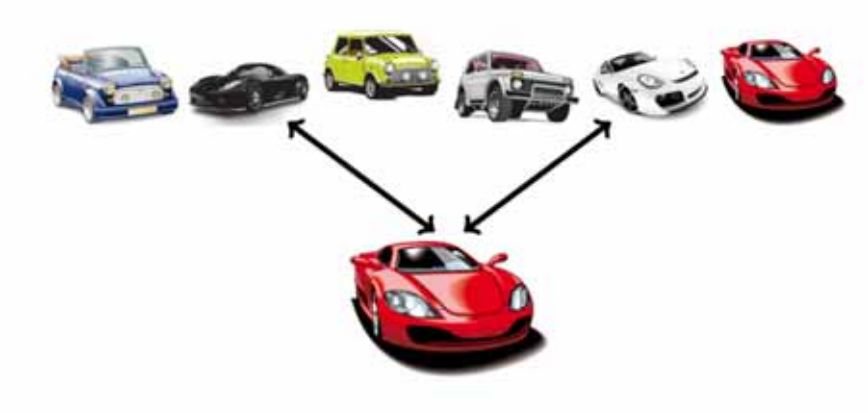
Poupart, P., Vlassis, N., Hoey, J. & Regan, K. An analytic solution to discrete Bayesian reinforcement learning. Proceedings of the 23rd international conference on Machine learning, 2006. ACM, 697-704.

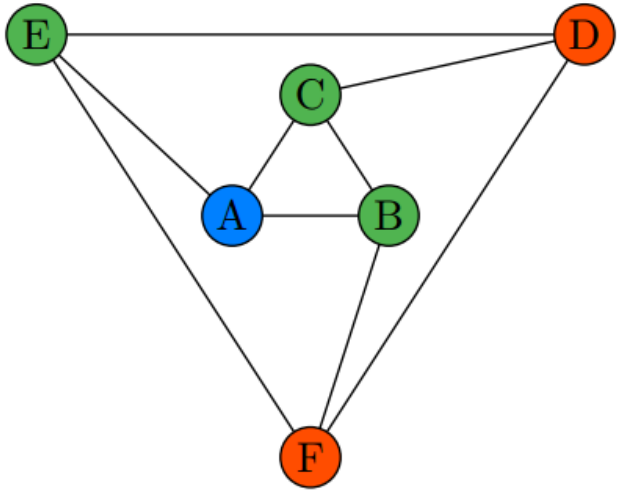
- Clinical trials: potential treatments for a disease to select from new patients or patient category at each round, see:

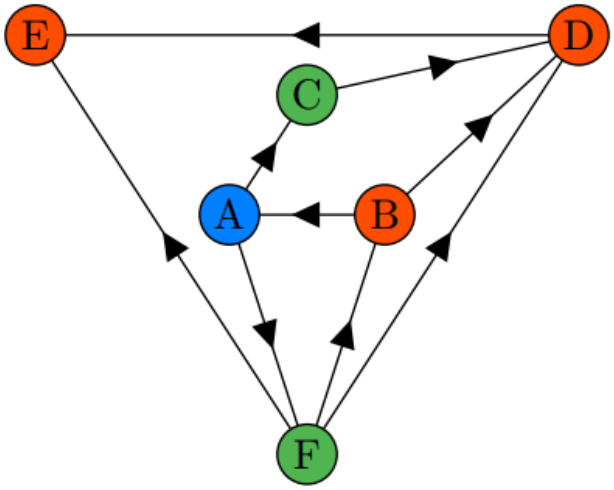
W. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Bulletin of the American Mathematics Society, vol. 25, pp. 285–294, 1933.

- Games: Different moves at each round, e.g. GO
- Adaptive routing: finding alternative paths, also finding alternative roads for driving from A to B
- Advertisement placements: selection of an ad to display at the Webpage out of a finite set which can vary over time, for each new Web page visitor

Graph Bandits -> undirected graph/directed graph







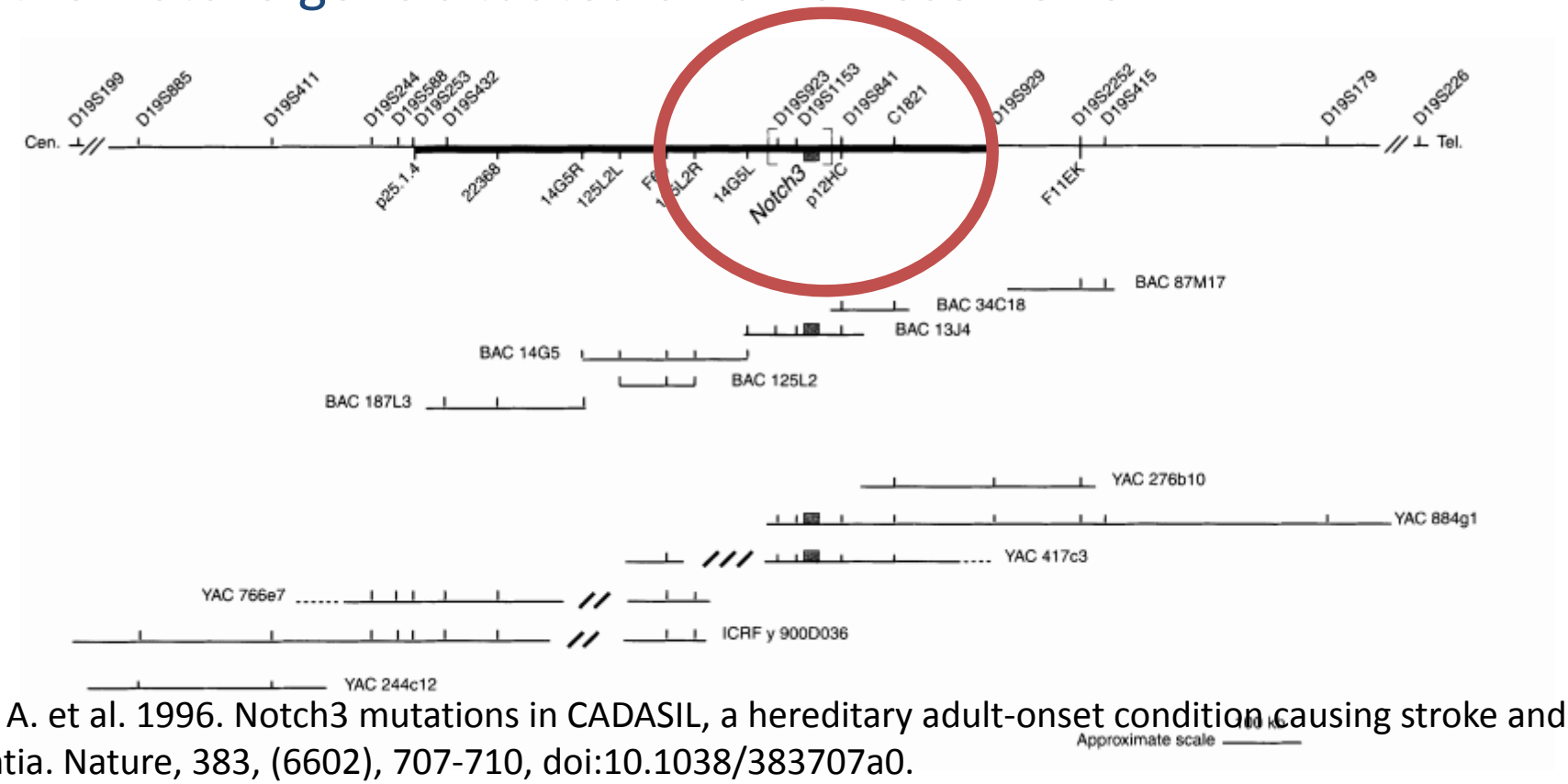
Some Statistics on rare diseases (orphan diseases)

- 7,000 + different types - more being discovered every day
- >10% of the world population is suffering (if all of the people with rare diseases lived in one country, it would be the world's 3rd most populous country)
- 80% of rare diseases are genetic, so are present throughout a person's lifetime, even if symptoms do not immediately appear
- >50% of the people affected by rare diseases are children
- Are responsible for 35% of deaths in the first year of life
- The prevalence distribution is skewed – 80% of all rare disease patients are affected by 350 rare diseases
- >50% of rare diseases do not have a disease specific foundation supporting or researching their rare disease

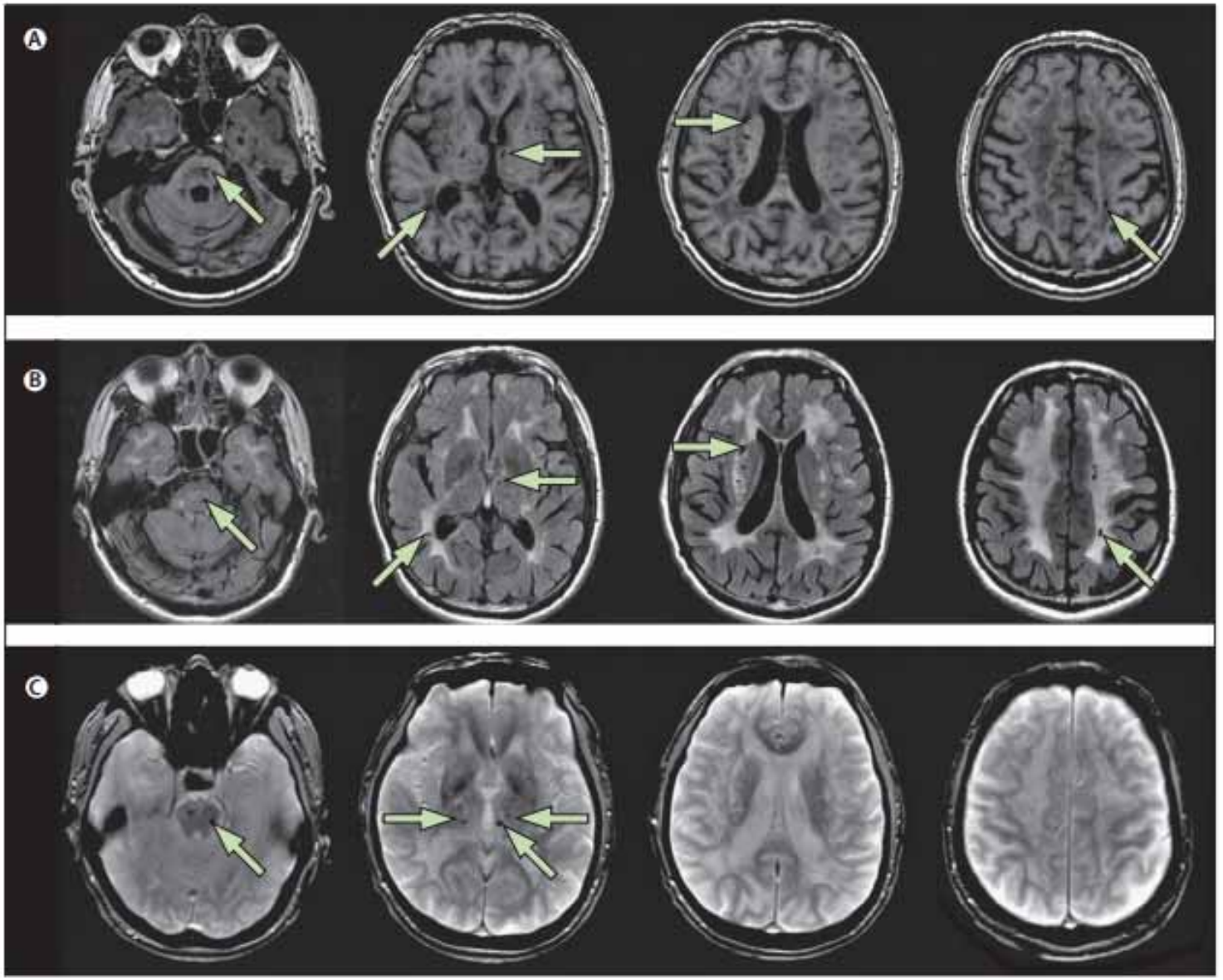
https://globalgenes.org/rare-diseases-facts-statistics/
https://www.hon.ch/HONselect/RareDiseases/

Example Rare Disease: CADASIL

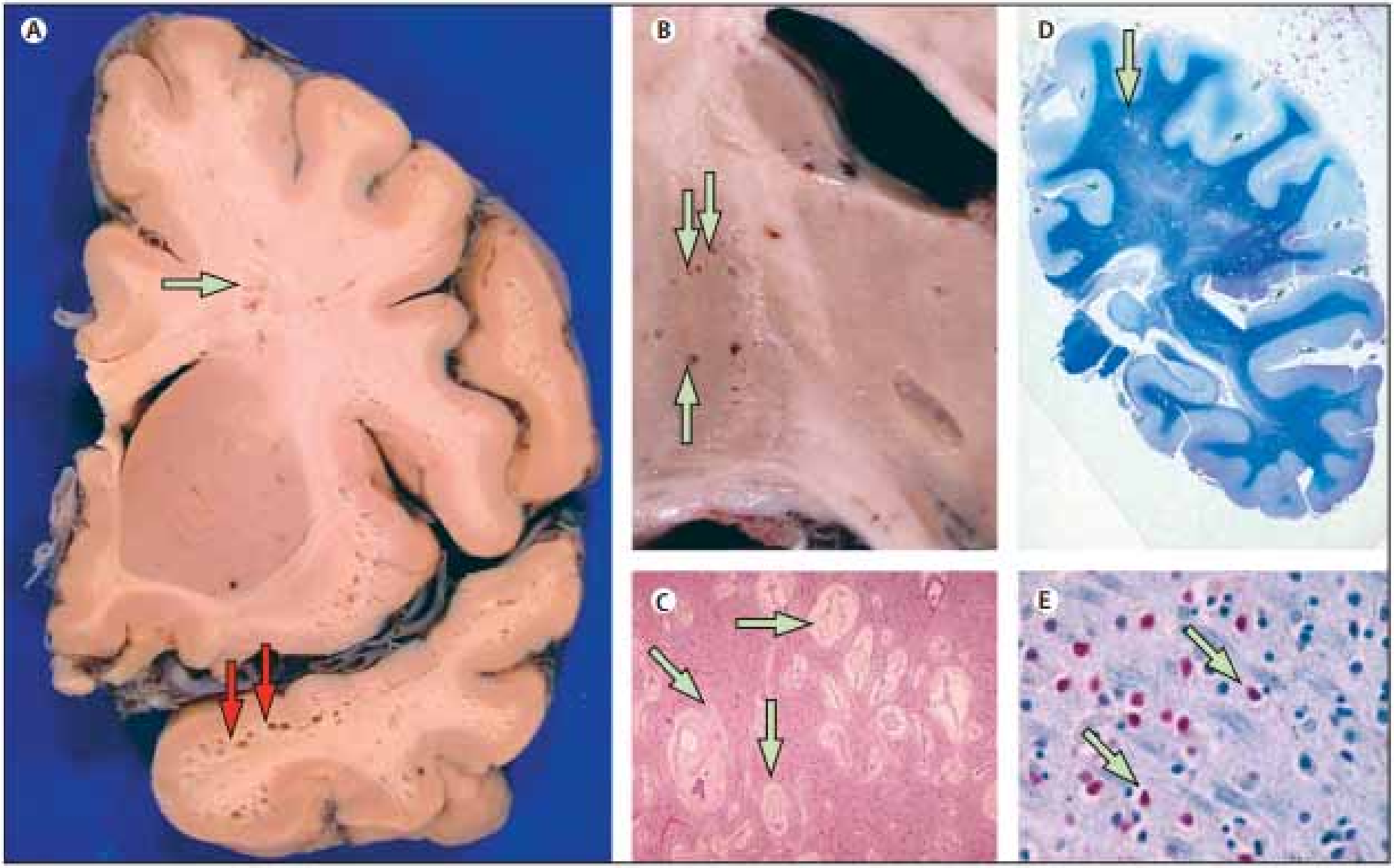
- Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
- is a hereditary disease affecting all the small cerebral arteries. It causes subcortical infarcts and damages the white matter (leukoencephalopathy) and it is due to various mutations of the Notch3 gene situated on chromosome 19:



Joutel, A. et al. 1996. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature, 383, (6602), 707-710, doi:10.1038/383707a0.

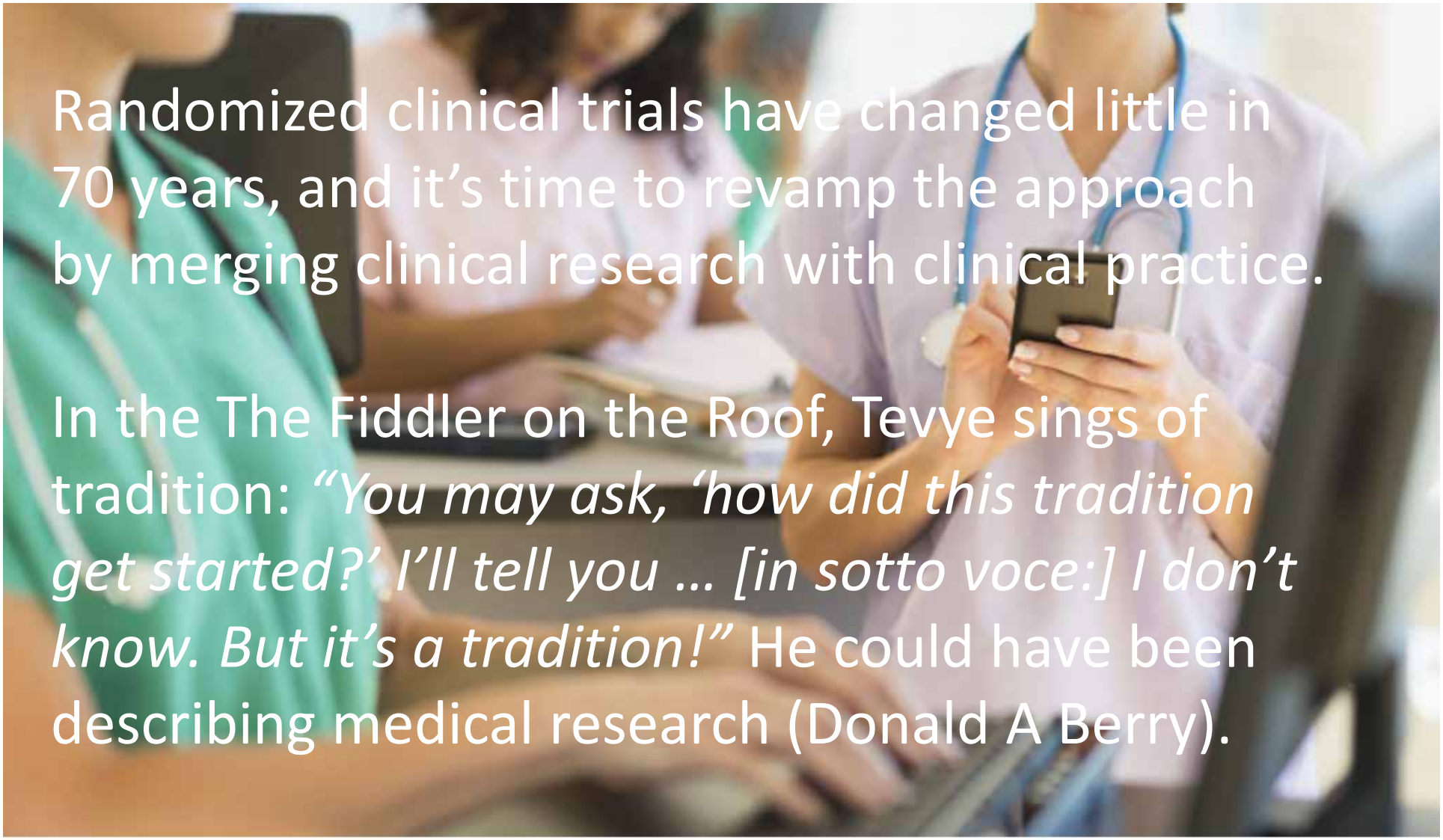


Chabriat, H., Joutel, A., Dichgans, M., Tournier-Lasserve, E. & Bousser, M.-G. 2009. CADASIL. The Lancet Neurology, 8, (7), 643-653, doi:http://dx.doi.org/10.1016/S1474-4422(09)70127-9.



Chabriat, H., Joutel, A., Dichgans, M., Tournier-Lasserve, E. & Bousser, M.-G. 2009. CADASIL. The Lancet Neurology, 8, (7), 643-653, doi:http://dx.doi.org/10.1016/S1474-4422(09)70127-9.

Generally: Long Tradition in medical research



http://fortune.com/2015/10/26/cancer-clinical-trial-belmont-report/

Example: Clinical Trials in rare diseases

Limitations of drug design for rare diseases due to:

- Lack of understanding of the underlying principles of the rare disease
  - Motivation: Research advances
- Unbalanced economic motivation (cost/benefit)
  - Motivation: Orphan Drug Act and other regulations
- Unavailability of # patients for standard trials
  - This is the true bottleneck!

Villar, S. S., Bowden, J. & Wason, J. 2015. Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. 199-215, doi:10.1214/14-STS504.

Why should Bandit Strategies be of help here?

- The goal of Standard Randomized Controlled Trials (RCT) are a controlled learning setting:
  - Control for Type I and Type II errors, dependent of trial size n_{RCT}
  - In the case if the patient population N is smaller than the trial size n_{RCT}: underpowered trial – problem!
- If we change the goal to
- *"learning sufficient - to treat N as effectively as possible",*
- then **bandit strategies** – optimal policy for max. the expected reward - are perfectly suited!

Kuleshov, V. & Precup, D. 2014. Algorithms for multi-armed bandit problems. *arXiv:1402.6028*.

Learning vs. Earning dilemma

- Learning → experimenting with all treatments
- Earning → selecting one treatment only, based on experimentation results
- **Question 1: How much learning is best – for an optimal treatment of N patients?**
- Suppose N patients with a rare disease:
  - Experimental Group E and control group C
  - e.g. control = response rate pc and little information about experimental group
- **Question 2: How many allocations of treatment to E are necessary (= how much experimentation?)**

Answer to Q2: Dynamic Programming

Richard Ernest BELLMAN (1920-1984)

DYNAMIC PROGRAMMING AND LAGRANGE MULTIPLIERS

By Richard Bellman

RAND CORPORATION, SANTA MONICA, CALIFORNIA

Communicated by Einar Hille, August 13, 1956

1. *Introduction.*—The purpose of this note is to indicate how a suitable combination of the classical method of the Lagrange multiplier and the functional-equation method of the theory of dynamic programming[1] can be used to solve numerically, and treat analytically, a variety of variational problems that cannot readily be treated by either method alone.

A series of applications of the method presented here will appear in further publications.

2. *Functional Equation Approach.*—Consider the problem of maximizing the function

$$F(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} g_i(x_i), \tag{2.1}$$

subject to the constraints

$$\begin{aligned} &(a) \quad \sum_{j=1}^{N} a_{ij}(x_j) \leq c_i, \qquad i = 1, 2, \ldots, M, \\ &(b) \quad x_i \geq 0, \end{aligned} \tag{2.2}$$

Bellman, R. 1956. Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences, 42, (10), 767-769.

- According to Whittle (1988) [1], we can now include a third treatment W with $p_W(N)$, so that for a given N it is indifferent between E and W
- If we know pW(N) then the answer to Q2 is: Allocate the experimental treatment as long as pw(N) > pc – otherwise switch to control group C

Restless Bandits: Activity Allocation in a Changing World

P. WHITTLE

Abstract

We consider a population of n projects which in general continue to evolve whether in operation or not (although by different rules). It is desired to choose the projects in operation at each instant of time so as to maximise the expected rate of reward, under a constraint upon the expected number of projects in operation. The Lagrange multiplier associated with this constraint defines an index which reduces to the Gittins index when projects not being operated are static. If one is constrained to operate m projects exactly then arguments are advanced to support the conjecture that, for m and n large in constant ratio, the policy of operating the m projects of largest current index is nearly optimal. The index is evaluated for some particular projects.

GITTINS INDEX; MULTI-ARMED BANDITS; SEQUENTIAL SCHEDULING; STIMULATING PRICES; INDEXABILITY

1. Introduction

The multi-armed bandit problem is a classic version of the problem of optimal allocation of activity under certainty. One can phrase it by saying that one has n projects, the state of project i being denoted by x_i (or by $x_i(t)$ if one wishes to emphasise its dependence on time, t). One can operate only one project at once: if one operates project i then one receives reward $g_i(x_i(t))$ in the time-interval $(t, t+1)$ and the transition $x_i(t) \rightarrow x_i(t+1))$ follows a Markov rule specific to project i. The unused projects neither yield reward nor change state; current states of all projects are known at any time. The problem is to so choose the project at each moment that the expected discounted reward over an infinite future is maximal.

The problem was proposed first during the Second World War, and had

- Start simple: $N = 1$ and $p_E \sim B(1, 1)$, what is the value of $p_W(1)$?

  $\max_{E,W}\{1/2, p_W(1)\} \rightarrow p_W(1) = 1/2 = 0.5$

  Then, if $p_C < 0.5$, treatment E is allocated.

- Now, consider $N = 2$ and $p_E \sim B(1, 1)$, what is the value of $p_W(2)$?

  $\max_{E,W}\{0.5(1 + 2/3) + 0.5(p_W(2)), 2p_W(2)\}$
  $\rightarrow p_W(2) = 5/9 = 0.5556.$

  Then, if $p_C < 0.5556$, treatment E is allocated.

Working recursively

- For $N = 2$, suppose that $p_C < 0.5556$, so treatment E is allocated to patient 1.

  Then, if a success is observed, $p_W(1) = 2/3$ treatment E is allocated to patient 2.

  Then, if a failure is observed $p_W(1) = 1/3$ treatment C is allocated to patient 2 if $p_C > 0.333$.

- Working recursively, these "Whittle" indices can be computed for any patient horizon N.

- For $N = 2$, suppose that $p_C < 0.5556$, so treatment E is allocated to patient 1.

  Then, if a success is observed, $p_W(1) = 2/3$ treatment E is allocated to patient 2.

  Then, if a failure is observed $p_W(1) = 1/3$ treatment C is allocated to patient 2 if $p_C > 0.333$.

- Working recursively, these "Whittle" indices can be computed for any patient horizon N.

Q2 is a Multi-armed Bandit Problem

- Suppose N patients with a rare life threatening disease.
- Two treatments available: control (C) and experimental (E)
- Suppose equipoise so that $p_E, p_C \sim B(1,1)$

Q2 How much experimentation? = How many allocations of treatments E and C until choosing one treatment for the rest of the population?

- The $p_W(N)$ can be computed independently for each treatment as the trial evolves.
- If 2 treatments available: C and E, then:

  Allocate C to the next patient only if $p_C(N) > p_E(N)$ (Ties broken at random)
- If multiple arms, then the rule is:

  Allocate to the next patient the treatment with the highest value of $p_K(N)$.

  For the multi-arm case the rule is a near-optimal heuristic.

- Bandit strategy: Is experimentation worth it for a small number N?
- Reconcile clinical trials and clinical practice
- Extensions should deal with randomization, delayed responses and uncertainty around N
- Bayesian bandits need Online-ML
- Bandits are a great source of inspirations and building blocks for solving manifold problems
- Future work: convex optimization, contextual, combinatorial, …

Berry, D. A. & Fristedt, B. 1985. Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability), Springer.

Thank you!

Sample Questions (1/3)

- Describe the clinical decision making process!
- Which type of graph is particularly useful for inference and learning?
- What is the key challenge in the application of graphical models for health informatics?
- What was Judea Pearl (1988) discussing in his paper, for which he received the Turing award?
- What main difficulties arise during breast cancer prognosis?
- What can be done to increase the robustness of prognostic cancer tests?
- Inference in Bayes Nets is NP-complete, but there are certain cases where it is tractable, which ones?

Sample Questions (2/3)

- Why do we want to apply ML to graphs?
- Describe typical ML tasks on the example of blood cancer cells!
- If you have a set of points – which similarity measures are useful?
- Why is graph comparison in the medical domain useful?
- Why is the Gromov-Hausdorff distance useful?
- What is the central goal of a generative probabilistic model?
- Describe the LDA-model and its application for topic modelling!

Sample Questions (2/3)

- Briefly describe the stochastic variational inference algorithms!
- What is the principle of a bandit?
- How does a multi-armed bandit (MAB) work?
- In which ways can a MAB represent knowledge?
- What is the main problem of a clinical trail – and maybe the main problem in clinical medicine?
- Why are rare diseases both important and relevant? Describe an example disease!
- What is the big problem in clinical trials for rare diseases?
- What did Richard Bellman (1956) describe with dynamic programming?
- Why are graph bandits a hot topic for ML research?

Solutions of the Quiz

- 1=this is a factor graph of an undirected graph – we have seen this in protein networks (refer to slide Nr. 70 in lecture 5). Factor graph is bipartite and has two types of nodes: Variables, which can be either evidence variables (when we know its value) or query variables (when the value is unknown and we want to predict the value); and factors, which define the relationship between variables in the graph. Each factor can be connected to many variables and comes with a factor function to define the relationship between these variables. For example, if a factor node is connected to two variables nodes A and B, a possible factor function could be imply(A,B), meaning that if the random variable A takes value 1, then so must the random variable B. Each factor function has a weight associated with it, which describes how much influence the factor has on its variables in relative terms. For more information please consult: http://deepdive.stanford.edu/inference
- 2= this is the decomposition of a tree, rooted at nodes into subtrees
- 3= an example for machine translation, Image credit to Kevin Gimpel, Carnegie Mellon University
- 4= the famous expectation-utility theory according to von Neumann and Morgenstern (1954): a decision-maker faced with risky (probabilistic) outcomes of different choices will behave as if he is maximizing the expected value of some function defined over the potential outcomes at some specified point in the future.
- 5= MYCIN –expert system that used early AI (rule-based) to identify bacteria causing severe infections, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight — the name derived from the antibiotics themselves, as many antibiotics have the suffix "-mycin".
- 6= metabolic and physical processes that determine the physiological and biochemical properties of a cell. These networks comprise the chemical reactions of metabolism, the metabolic pathways, as well as the regulatory interactions that guide these reactions.
- 7= With the sequencing of complete genomes, it is now possible to reconstruct the network of biochemical reactions in many organisms, from bacteria to human. Several of these networks are available online, e.g. Kyoto Encyclopedia of Genes and Genomes (KEGG), EcoCyc, BioCyc etc. Metabolic networks are powerful tools for studying and modelling metabolism.