

Andreas Holzinger



185.A83 Machine Learning for Health Informatics
2017S, VU, 2.0 h, 3.0 ECTS

Lecture 06 - Module 04 – Week 17 - 25.04.2017



Probabilistic Graphical Models

Part 2: From Bayesian Networks

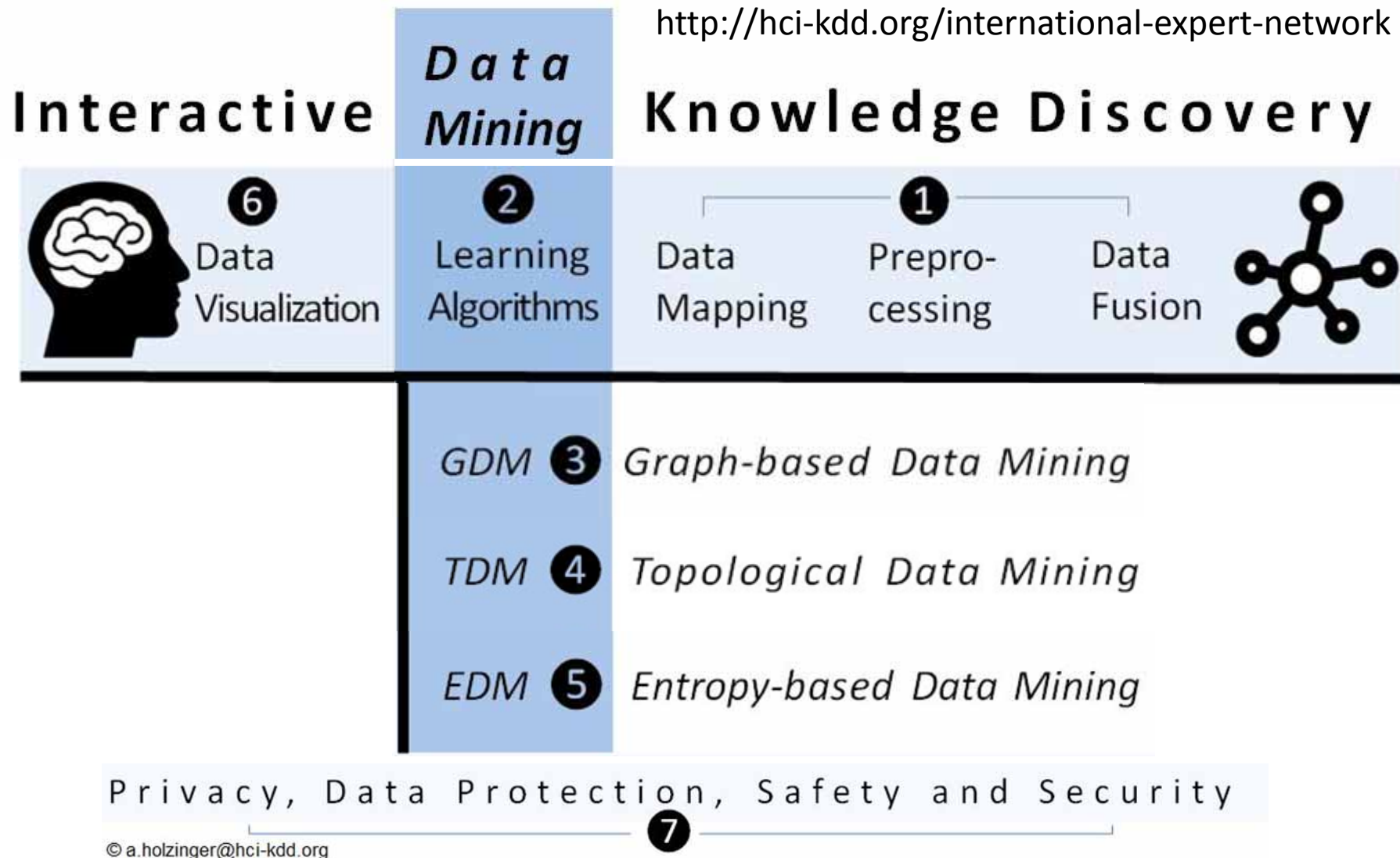
Probabilistic Topic Models

a.holzinger@hci-kdd.org

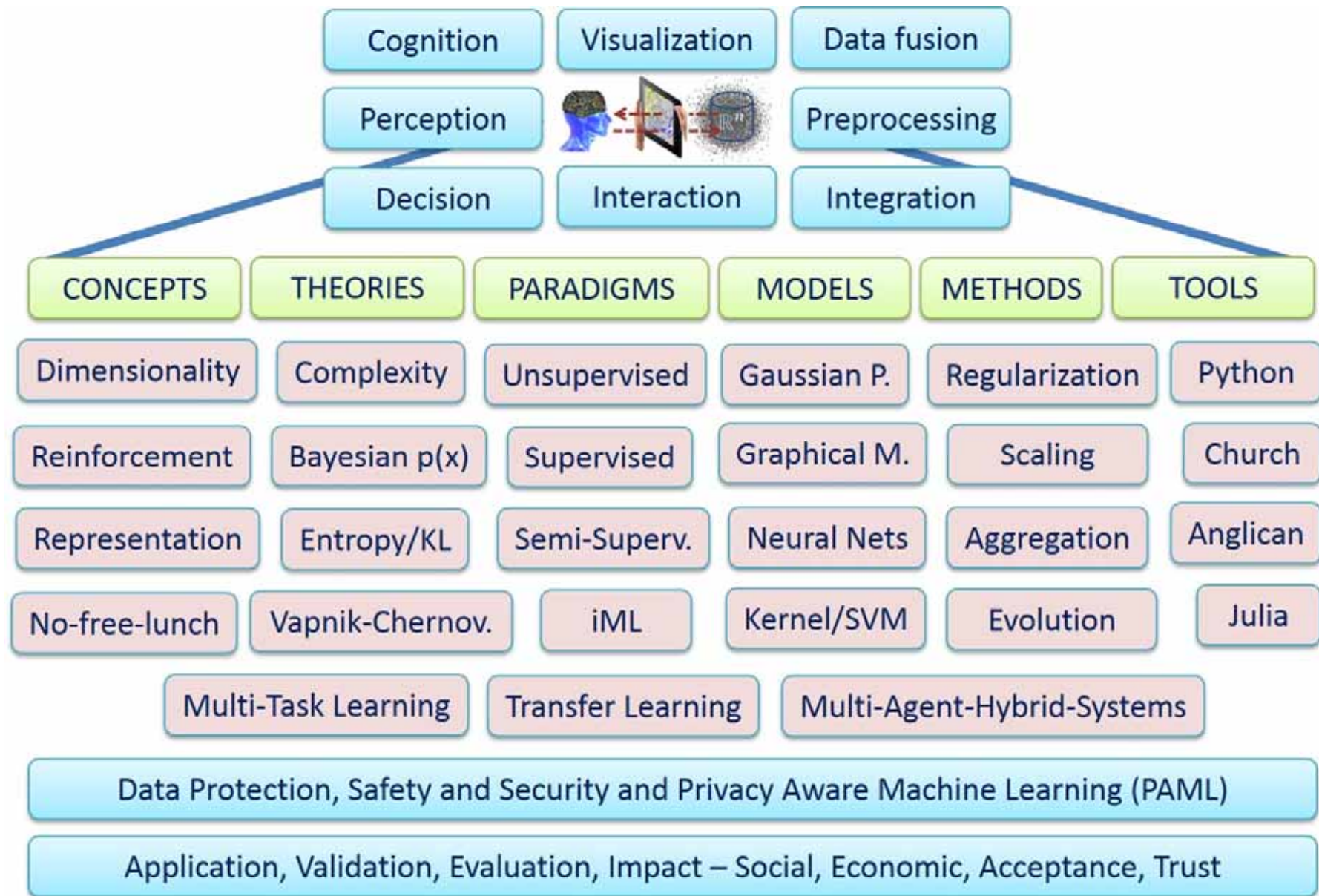
<http://hci-kdd.org/machine-learning-for-health-informatics-course>



<http://hci-kdd.org/international-expert-network>



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



Holzinger, A. 2016. Machine Learning for Health Informatics. In: LNCS 9605, pp. 1-24, doi:10.1007/978-3-319-50478-0_1.

- **01 Probabilistic Decision Making**
- **02 Probabilistic Topic Models**
- **03 Knowledge Representation in Net Medicine**
- **04 ML on Graphs Examples**
- **05 Digression: Similarity**
- **06 Graph Measures**
- **07 Point Clouds from Natural Images**



00 Reflection

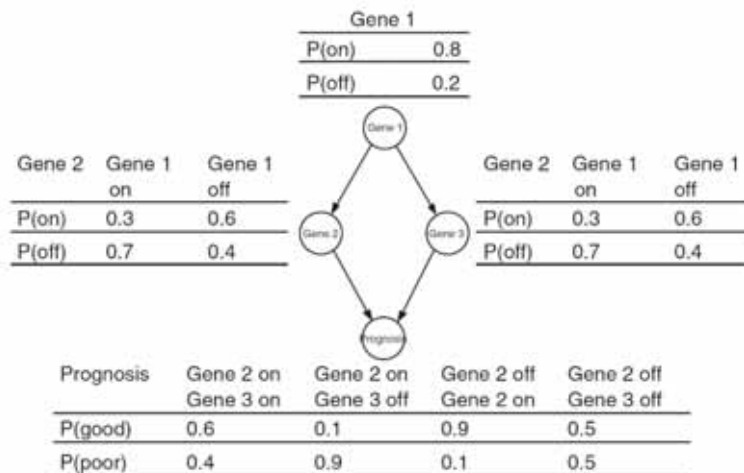
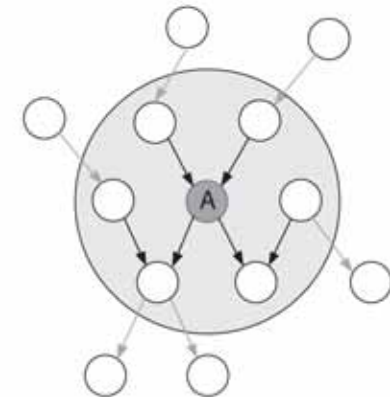
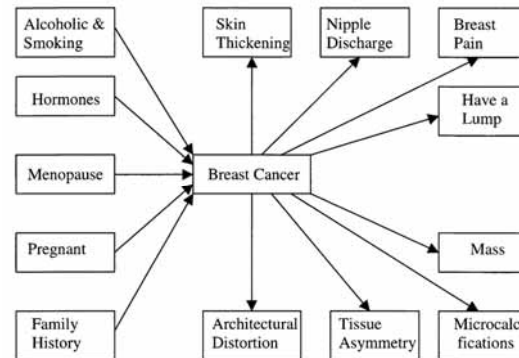
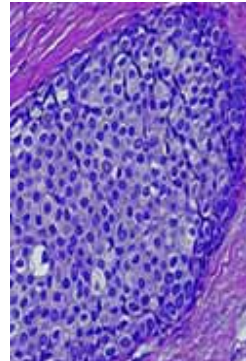
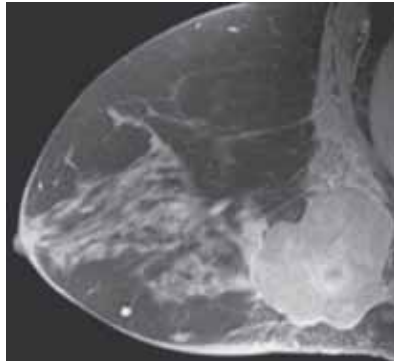
<http://smashinghub.com/beautiful-examples-of-shadow-photography.html>

- 1) **learn** from prior data
- 2) **extract** knowledge
- 2) **generalize**,
 - i.e. guessing where a probability mass function concentrates
- 4) fight the curse of **dimensionality**
- 5) **disentangle** underlying explanatory factors of data, i.e.
- 6) **understand** the data in the **context** of an application domain

$$\mathbb{E}[f] = \int f(z)p(z)dz$$



$$\hat{f} = \frac{1}{L} \sum_{l=1}^L f(z^{(l)})$$



$$P(z_1, \dots, z_n) = \prod_{i=1}^n P(z_i | pa(z_i))$$

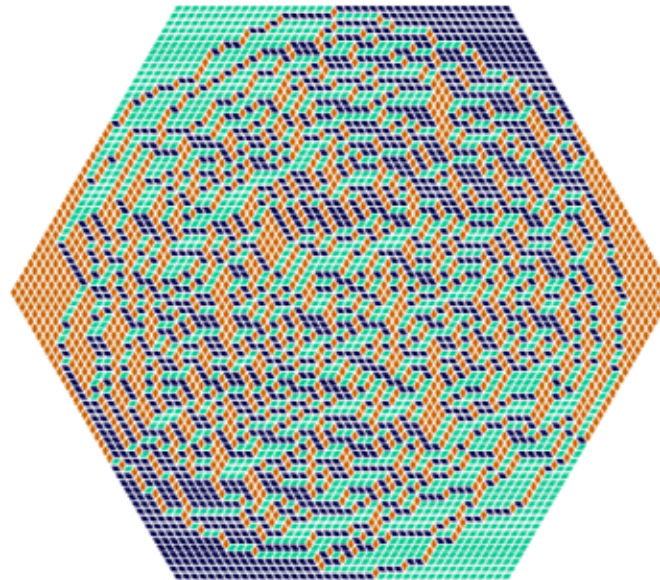
Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y. & Moor, B. D. (2006) Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics*, 22, 14, 184-190.

- 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

$$\mathbb{E}[f] = \int f(z)p(z)dz$$



$$\hat{f} = \frac{1}{L} \sum_{l=1}^L f(z^{(l)})$$

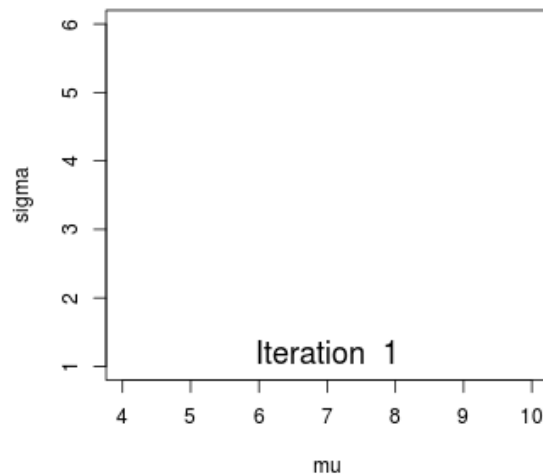


```

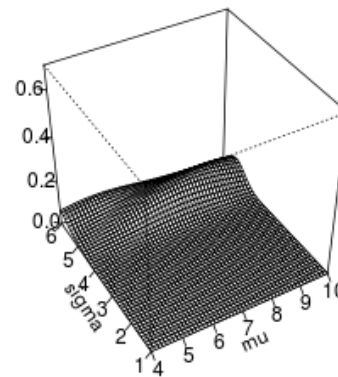
Compute  $a_i := \sum_j J_{ij}x_j$ 
Draw  $u$  from Uniform(0, 1)
If  $u < 1/(1 + e^{-2a_i})$ 
   $x_i := +1$ 
Else
   $x_i := -1$ 
    
```



Markov chains

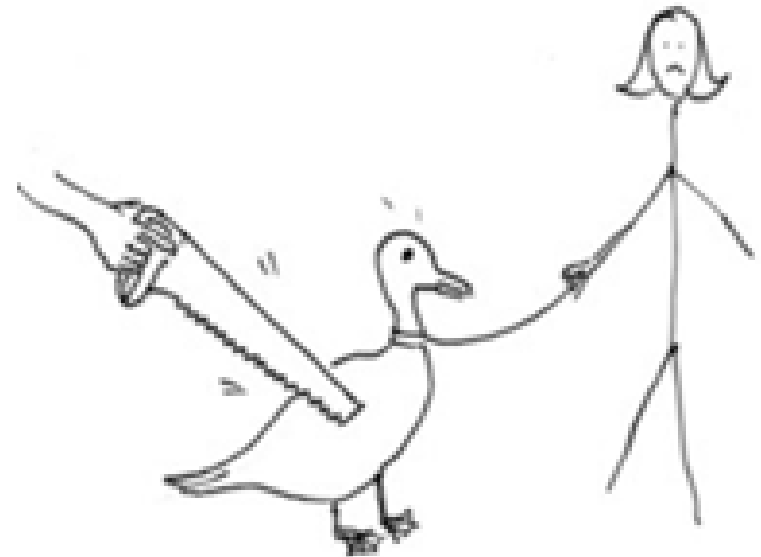
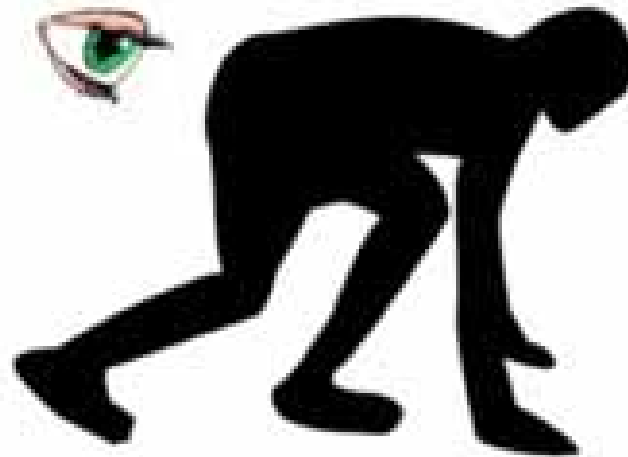
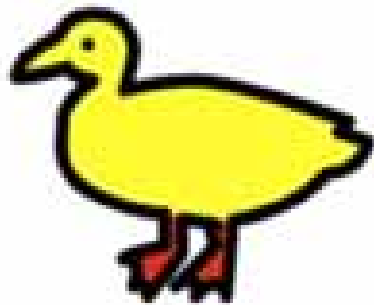


Posterior density



Propp, J. G. & Wilson, D. B.
1996. Exact sampling with
coupled Markov chains and
applications to statistical
mechanics. Random
structures and Algorithms,
9, (1-2), 223-252.

“I saw her duck”



Radiologischer Befund		angelegt am 06.05.2006/20:26 geschr. von [REDACTED] gedruckt am 17.11.2006/08:24 Anfo: NCHIN
Kurzanamnese:	St.p. SHT	
Fragestellung:	-	
Untersuchung:	Thorax eine Ebene liegend [REDACTED]	
SB		
Bewegungsartefakte. Zustand nach Schädelhirntrauma.		
Das Cor in der Größennorm, keine akuten Stauungszeichen. Fragliches Infiltrat parahilär li. im UF, RW-Erguss li.		
Zustand nach Anlage eines ET, die Spitze ca. 5cm cranial der Bifurkation, lieg. MS, orthotop positioniert. ZVK über re., die Spitze in Proj. auf die VCS. Kein Hinweis auf Pneumothorax. Der re. Rezessus frei.		
Mit kollegialen Grüßen		
[REDACTED]		
*** Elektronische Freigabe durch [REDACTED] am 09.05.2006 ***		

Special Words
Language Mix
Abbreviations
Errors ...

Holzinger, A., Geierhofer, R. & Errath, M. 2007. Semantische Informationsextraktion in medizinischen Informationssystemen. *Informatik Spektrum*, 30, (2), 69-78.

- HWI =
 - Harnwegsinfekt
 - Hinterwandinfarkt
 - Hinterwandischämie
 - Hakenwurminfektion
 - Halswirbelimmobilisation
 - Hip Waist Index
 - Height-Width Index
 - Heart-Work Index
 - Hemodynamically weighted imaging
 - High Water Intake
 - Hot water irrigation
 - Hepatitic weight index
 - Häufig wechselnder Intimpartner

- Leitung = Nervenleitung, Abteilungsleitung, Stromleitung, Wasserleitung, Harnleitung, Ableitung, Vereinsleitung ☺...



01 Probabilistic Decision Making

Laplace, P.-S. 1781. Mémoire sur les probabilités. *Mémoires de l'Académie Royale des sciences de Paris*, 1778, 227-332.



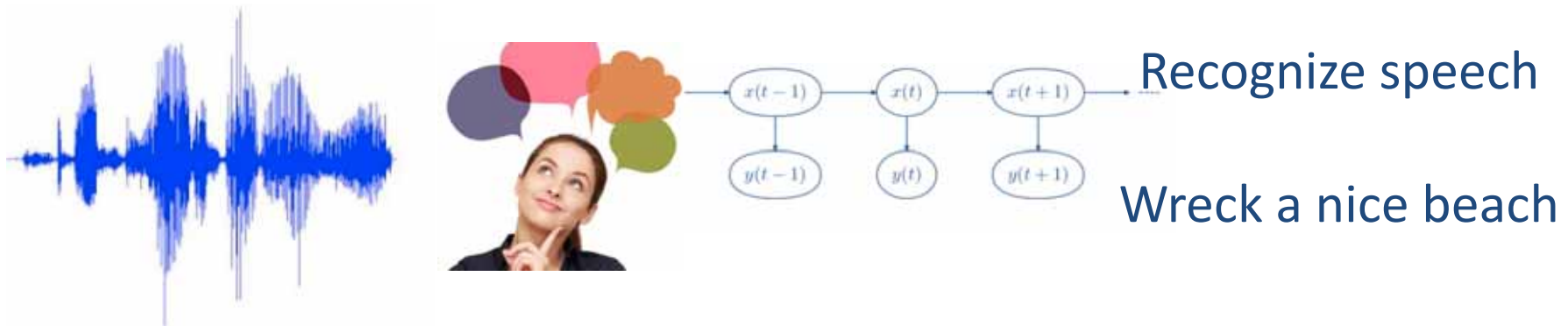
Medical action ...

is permanent
decision making under
uncertainty ...

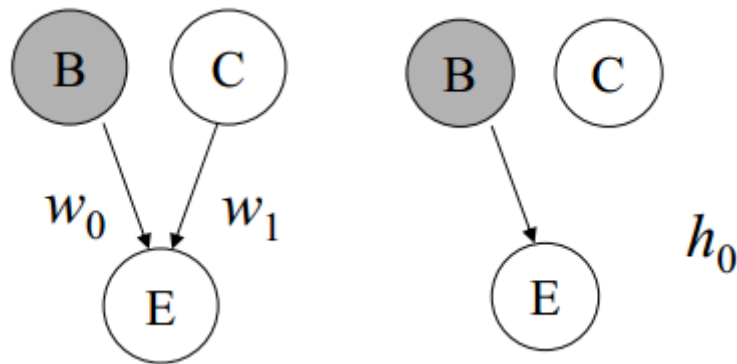
- Example 1: Inverse Probability
- Example 2: Diagnosis
- Example 3: Language understanding:

$$p(h|d) \propto p(\mathcal{D}|\theta) * p(h)$$

$$P(words|sounds) \propto P(sounds|words) * P(words)$$



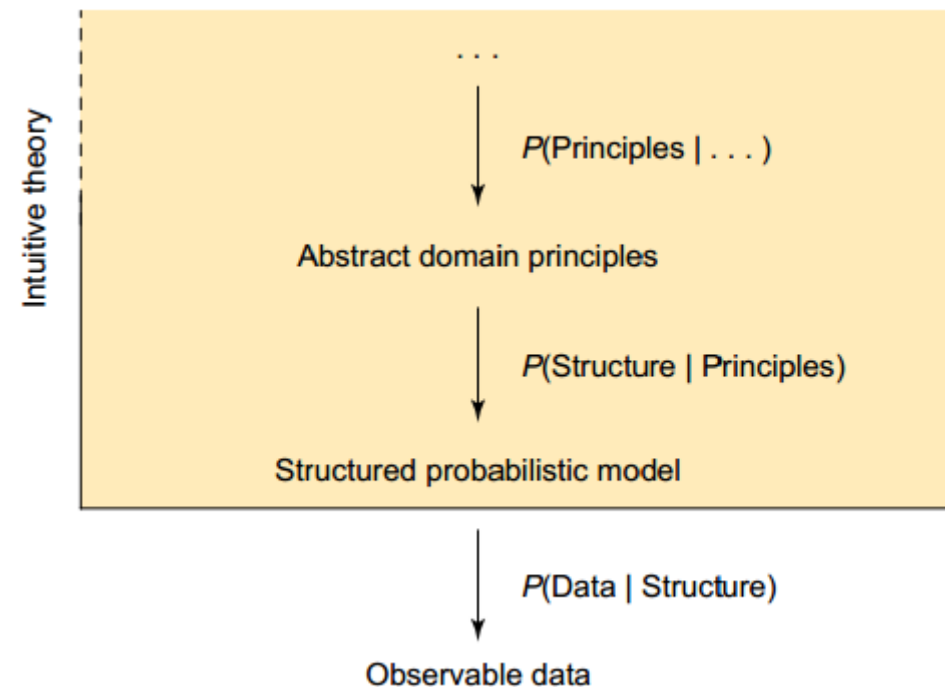
- Learning ensures that new observations (d) match our previous hypotheses (h)

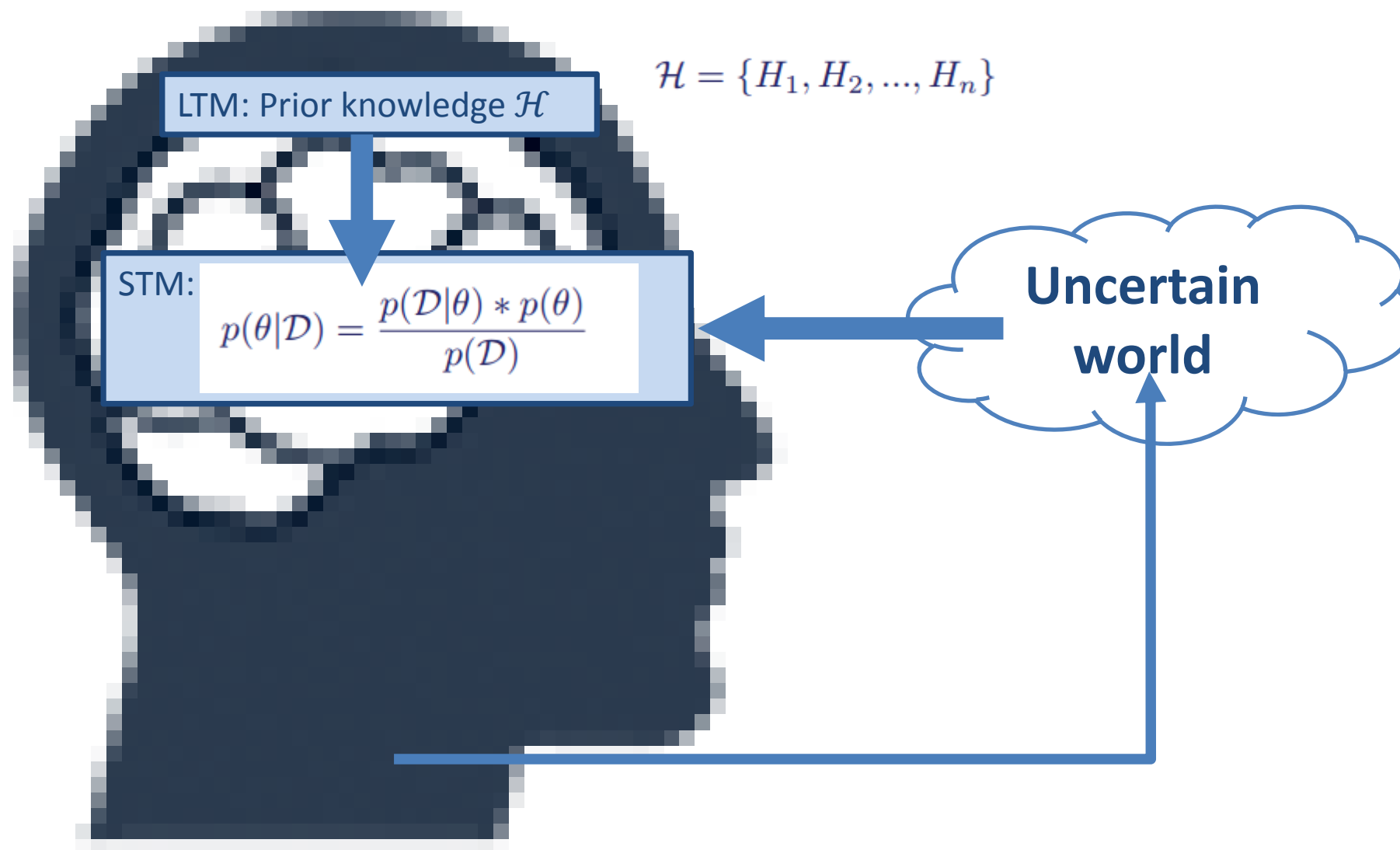


- Visual perception, language understanding, motor learning, associative learning, categorization, concept learning, reasoning, causal inference, ...
- Learning concepts from (few!) examples
- Learning and applying intuitive theories (balancing complexity vs. fit optimality)

- Similarity
- Representativeness and evidential support
- Causal judgement
- Coincidences and causal discovery
- Diagnostic inference
- Predicting the future

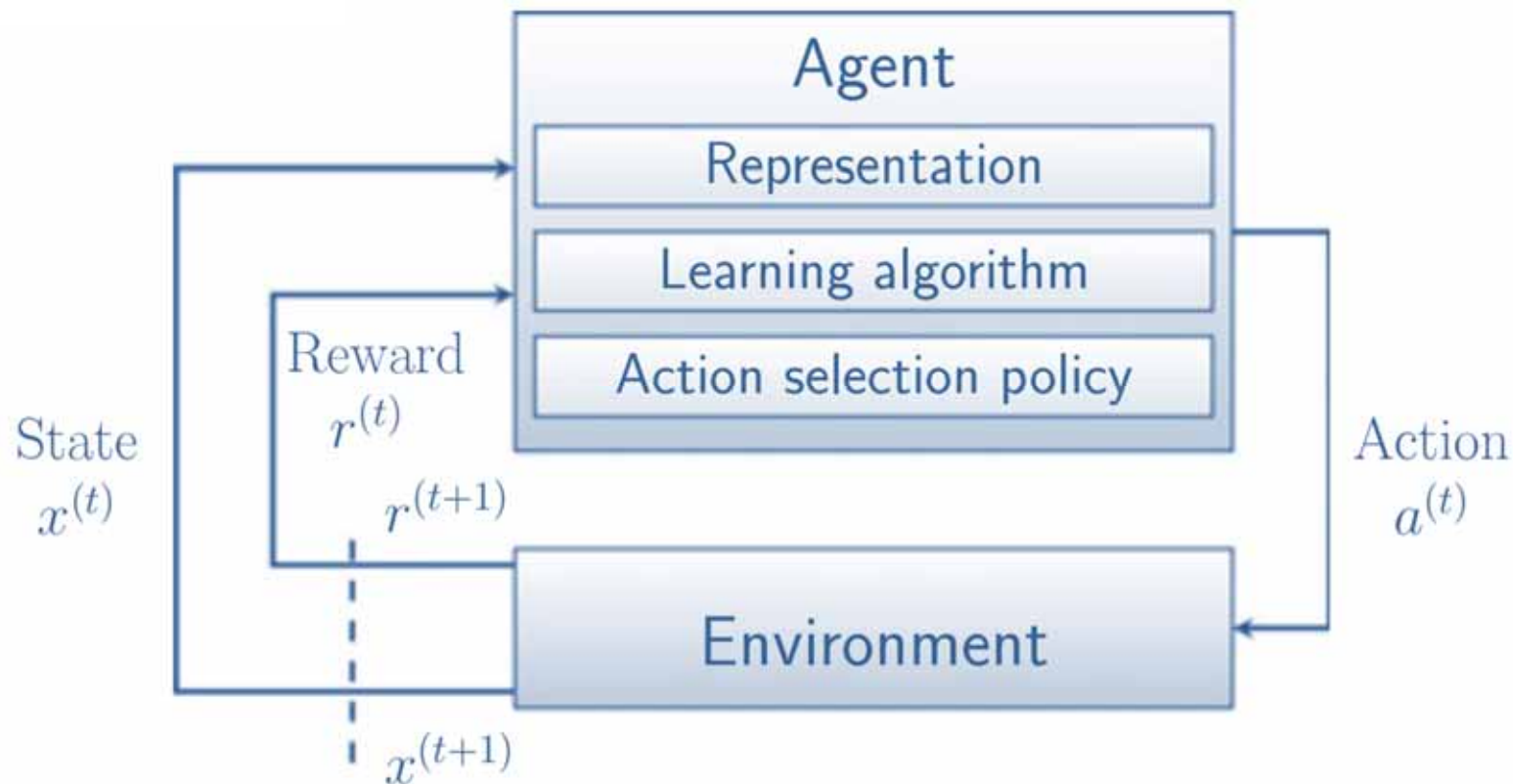
Tenenbaum, J. B., Griffiths, T. L. & Kemp, C. 2006. Theory-based Bayesian models of inductive learning and reasoning. Trends in cognitive sciences, 10, (7), 309-318.



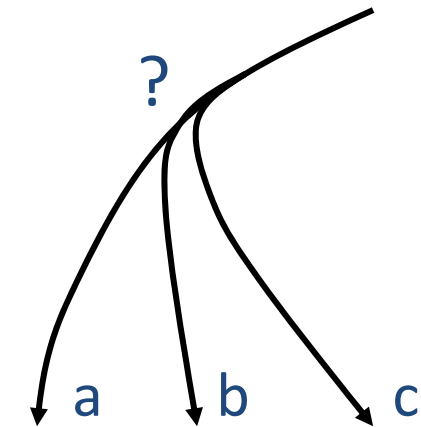


```
for  $t = 1, \dots, n$  do  
  The agent perceives state  $s_t$   
  The agent performs action  $a_t$   
  The environment evolves to  $s_{t+1}$   
  The agent receives reward  $r_t$   
end for
```

Intelligent behavior arises from the actions of an individual seeking to **maximize its received reward** signals in a **complex and changing world**



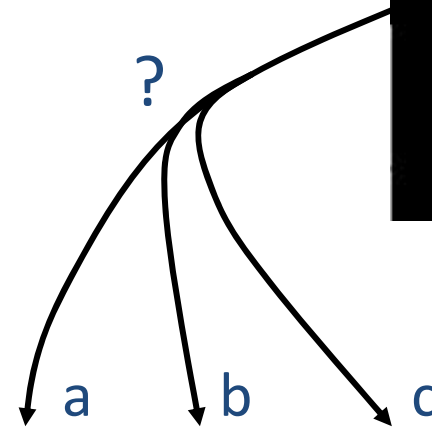
Sutton, R. S. & Barto, A. G. 1998. Reinforcement learning: An introduction, Cambridge MIT press

 $\{a, b, c\}$

→ decision that is
best for worst case

Non-deterministic model

~ Adversarial search

 $\{a(p_a), b(p_b), c(p_c)\}$

→ **decision that maximizes
expected utility value**

Probabilistic model



For a single decision variable an agent can select $D = d$ for any $d \in \text{dom}(D)$.

The expected utility of decision $D = d$ is



<http://www.eoht.info/page/Oskar+Morgenstern>

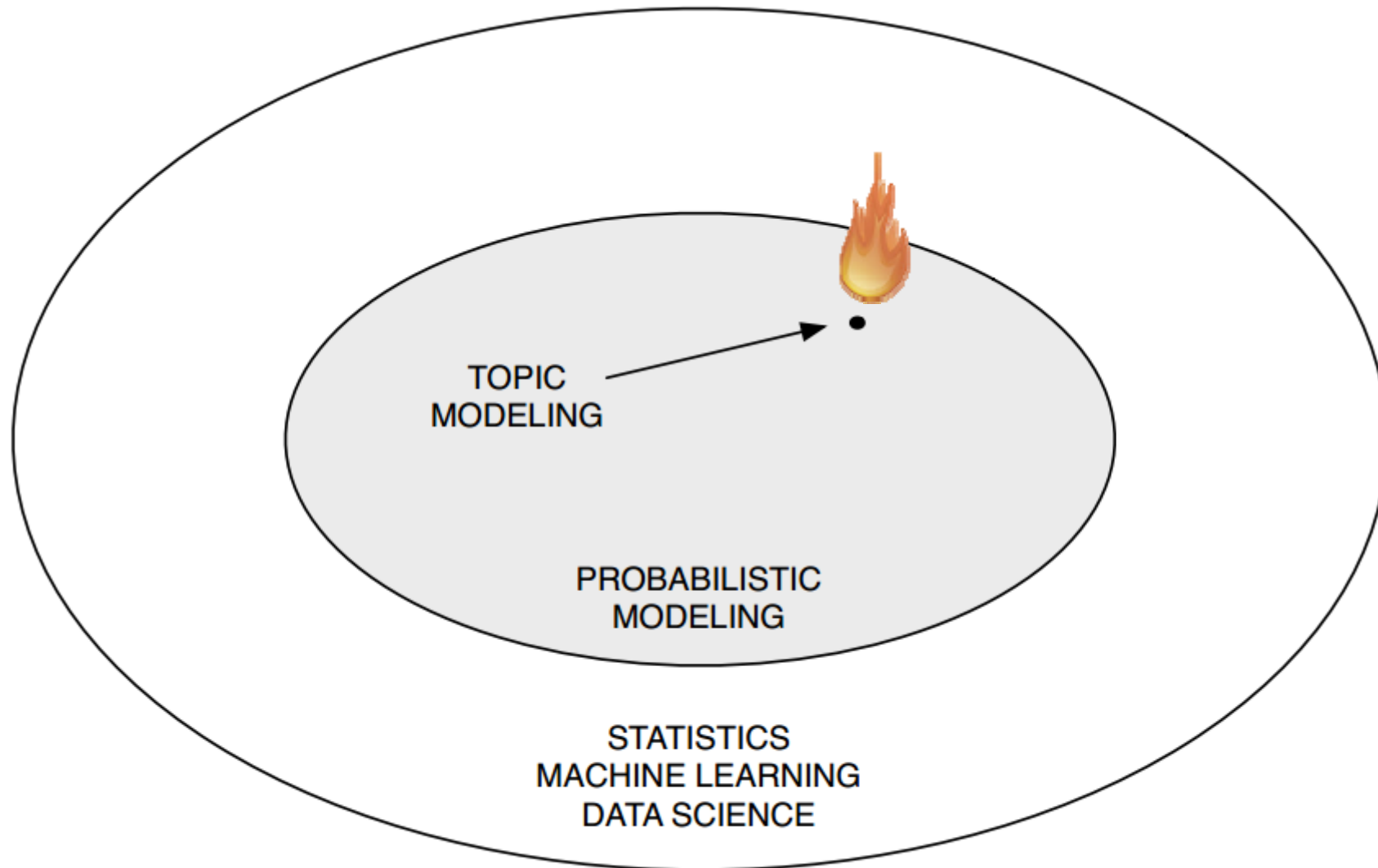
$$E(U \mid d) = \sum_{x_1, \dots, x_n} P(x_1, \dots, x_n \mid d) U(x_1, \dots, x_n, d)$$

An optimal single decision is the decision $D = d_{\max}$ whose expected utility is maximal:

$$d_{\max} = \arg \max_{d \in \text{dom}(D)} E(U \mid d)$$

Von Neumann, J. & Morgenstern, O. 1947. Theory of games and economic behavior, Princeton university press.

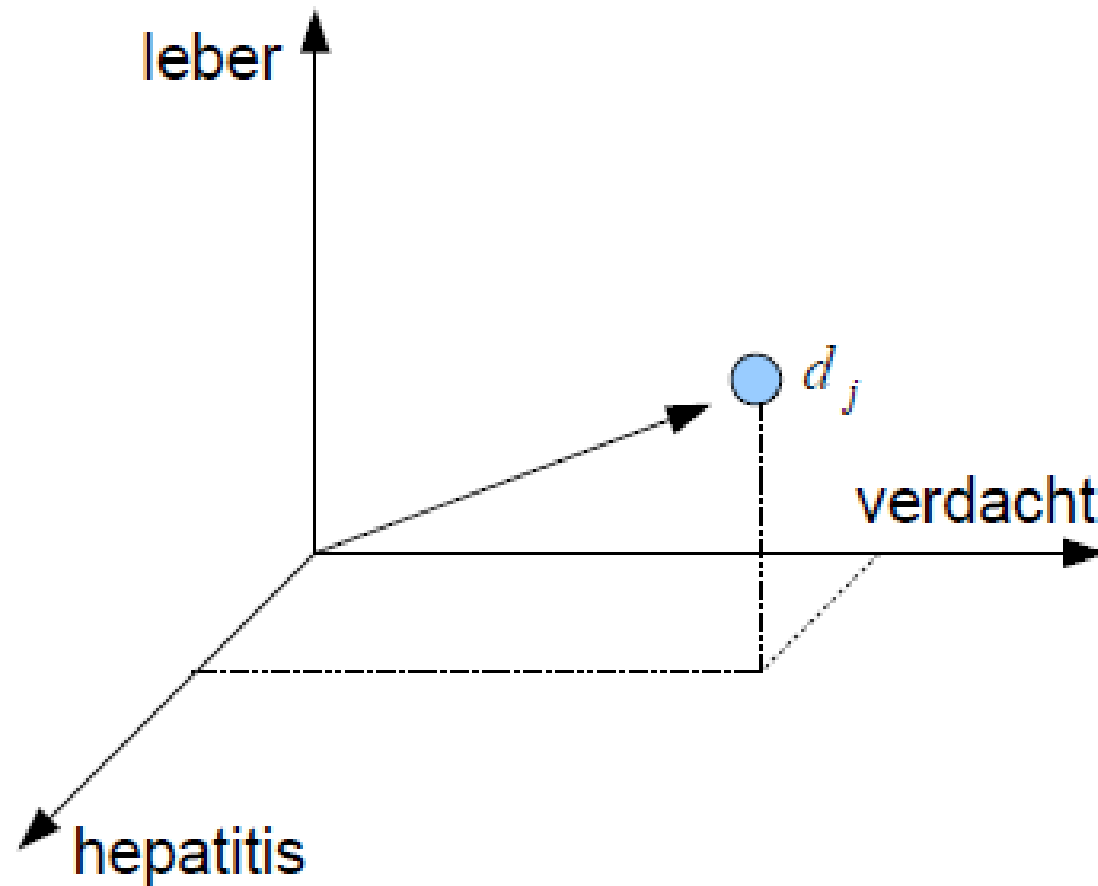
02 Probabilistic Topic Models

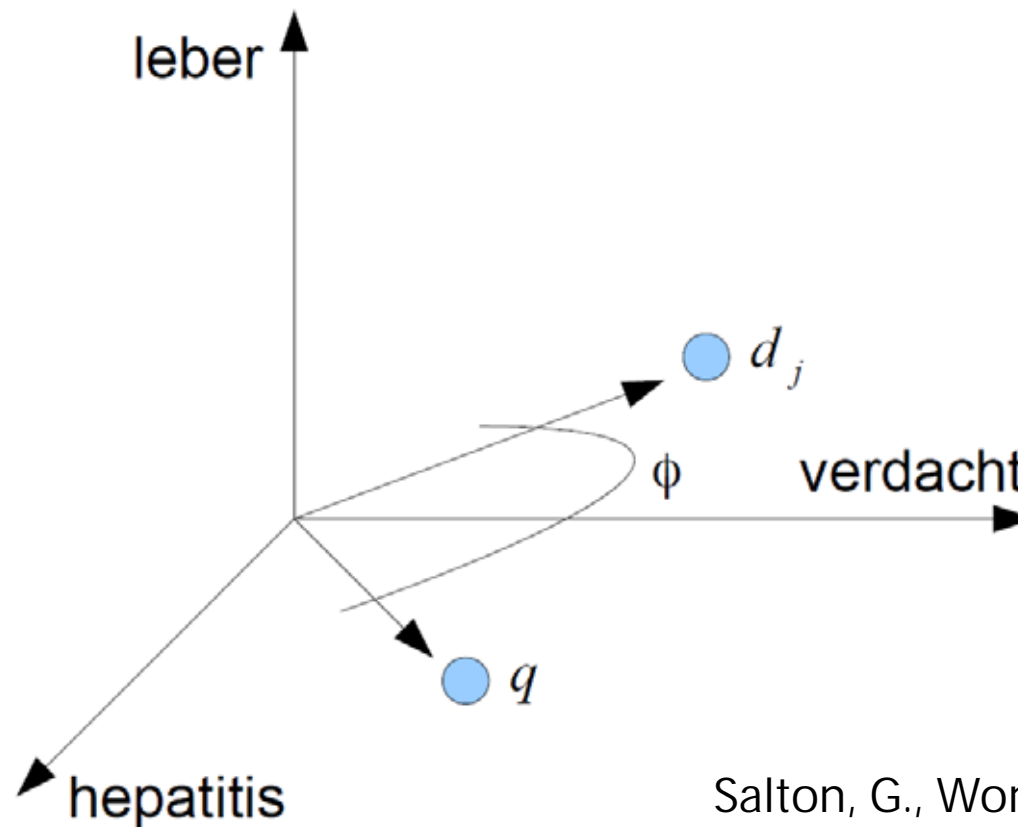


		Datum	Zeit	OE	Kurzbez.	VMA	Status	Klass.	erg. Bez.	Dokum...	Fall
▼	2010018065	17.11.1953									
▼	2010018065 ambulanter Fall	13.01.2010			MKKARDIO MK KardioAmb						
▼	Leistungen (KAL, RAD, Therap										
	EKG (12 Ableitungen)	13.01.2010	08:00	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2010018065
	RR-Intervall-Untersuchung	13.01.2010	08:00	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2010018065
	Schrittmacherkontrolle	13.01.2010	08:00	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2010018065
▼	2010002197 ambulanter Fall	04.01.2010			CKTRANSFCK Transpl.						
▶	Diagnosen Gesamt (3)										
▼	2009494995 stationärer Fall	20.12.2009			MEDANGIO Med Angio						
▶	Diagnosen Gesamt (14)										
▼	Leistungen (KAL, RAD, Therap										
	Becken-u. Beinarteriografie	22.12.2009	16:36	RKVIRADB RKVI Raum B	STANMELI	OK		PTA			2009494995
	Laufbandergometer	21.12.2009	08:30	MKANGIO MK AngioAmb	SPARANDR	OK					2009494995
	Erstuntersuchung/Status	21.12.2009	08:30	MKANGIO MK AngioAmb	SPARANDR	OK					2009494995
▼	2009453621 stationärer Fall	17.11.2009			CKTXIMC CK TX IMC						
▶	Diagnosen Gesamt (12)										
▼	Leistungen (MEL) (2)										
	Physioth. i.R.1 stat. Aufenth	23.11.2009	08:05	CKPHYSIO CK Physio	BEITWALT	OK					2009453621
	Organbiop., Bildwandlerge	17.11.2009	08:12	CKTXOP CK TX OP	SCHWMICH	OK		Organbiopsie - Bildwandler			2009453621
▶	Leistungen (KAL, RAD, Therap										
▼	2009431136 ambulanter Fall	29.10.2009			MKKARDIO MK KardioAmb						
▼	Leistungen (KAL, RAD, Therap										
	Schrittmacherkontrolle	29.10.2009	09:15	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2009431136
	RR-Intervall-Untersuchung	29.10.2009	09:15	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2009431136
	EKG (12 Ableitungen)	29.10.2009	09:15	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2009431136
	Fotodokumentation, Videoc	29.10.2009	09:15	MKKARDIO MK KardioAmb	DUSLTIMO	OK					2009431136
▼	2009378733 ambulanter Fall	16.09.2009			MKNEPHR(MK NephroAmb						
▶	Diagnosen Gesamt (8)										
▼	Leistungen (KAL, RAD, Therap										
	Blutdruck: Langzeit (24 Stur	17.09.2009	10:59	MKNEPHR(MK NephroAmb	RUDRHELM	OK		ab			2009378733
	Blutdruck: Langzeit (24 Stur	16.09.2009	12:02	MKNEPHR(MK NephroAmb	RUDRHELM	OK		an			2009378733
▼	2009187546 stationärer Fall	21.04.2009			CKGMIÜ CK GM IÜ						
▶	Diagnosen Gesamt (5)										
▼	Leistungen (KAL, RAD, Therap										
	Fotodokumentation, Videoc	29.04.2009	08:49	MKKARDIO MK KardioAmb	PITTHEID	OK					2009187546
	EKG (12 Ableitungen)	29.04.2009	08:49	MKKARDIO MK KardioAmb	PITTHEID	OK					2009187546
	RR-Intervall-Untersuchung	29.04.2009	08:49	MKKARDIO MK KardioAmb	PITTHEID	OK					2009187546
	Schrittmacherkontrolle	29.04.2009	08:49	MKKARDIO MK KardioAmb	PITTHEID	OK					2009187546
	EKG (12 Ableitungen)	23.04.2009	10:43	MKKARDIO MK KardioAmb	KOBEINGR	OK					2009187546
	RR-Intervall-Untersuchung	23.04.2009	10:43	MKKARDIO MK KardioAmb	KOBEINGR	OK					2009187546
	Konsil FA	21.04.2009	10:22	NKKONS NK FA Konsil	LANNMICH	OK					2009187546

- $D = \langle d_1, d_2, \dots, d_n \rangle$
- $d_i = t_1, t_2, \dots, t_k$
- $w_{i,j} = \begin{cases} 1, & t_i \in d_j \\ 0, & t_i \notin d_j \end{cases} \rightarrow d_j = (0, 1, 1, 0, 1, \dots, 1)^T$
- $w_{i,j} = \begin{cases} (1 + \log f_{i,j}) * \log \frac{N}{n_i}, & \text{if } f_{i,j} > 0 \\ 0 & \text{otherwise} \end{cases}$

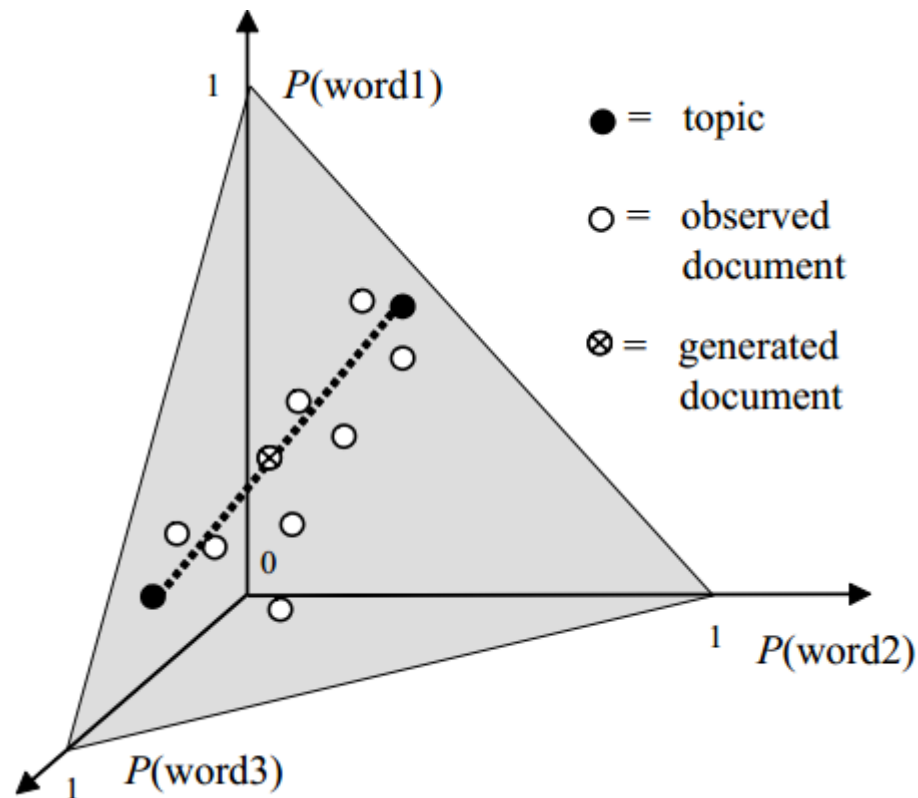
$$D_{m \times n} = \left\{ \begin{array}{ccccc} w_{1,1} & w_{1,2} & \cdots & w_{1,n-1} & w_{1,n} \\ w_{2,1} & w_{2,2} & & w_{2,n-1} & w_{2,n} \\ \vdots & & \ddots & & \vdots \\ w_{m-1,1} & w_{m-1,2} & & w_{m-1,n-1} & w_{m-1,n} \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n-1} & w_{m,n} \end{array} \right\}$$





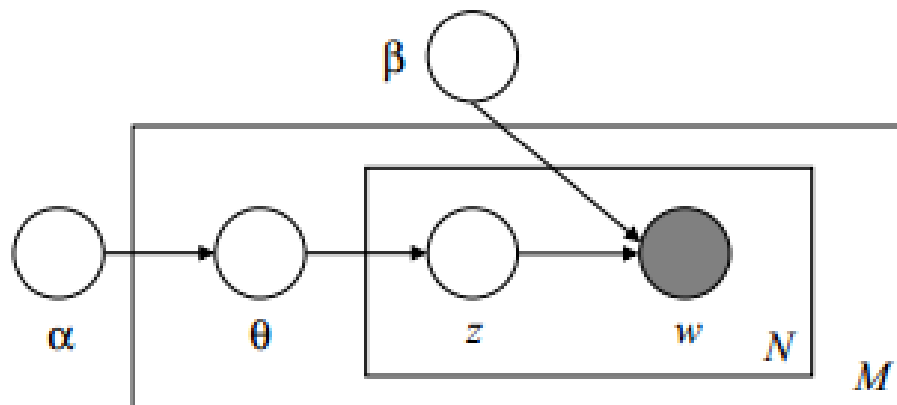
$$\cos(\phi) = \frac{q \cdot d_j}{\|q\| \|d_j\|}$$

Salton, G., Wong, A. & Yang, C. S. 1975.
Vector-Space Model for automatic
indexing. *Communications of the ACM*, 18,
(11), 613-620.



- Documents = categorical distributions over a large space of predefined vocabulary
- Topics = categorical distributions
- Generative model = each document can be seen as a convex combination of the topic distributions

Teh, Y. W., Jordan, M. I., Beal, M. J. & Blei, D. M. 2006. Hierarchical dirichlet processes. Journal of the american statistical association, 101, (476), 1566-1581.

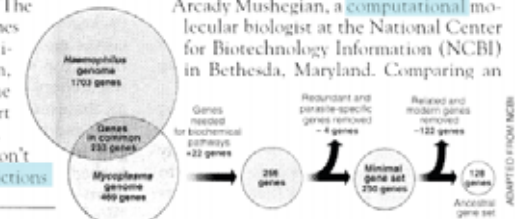


Seeking Life's Bare (Genetic) Necessities

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.

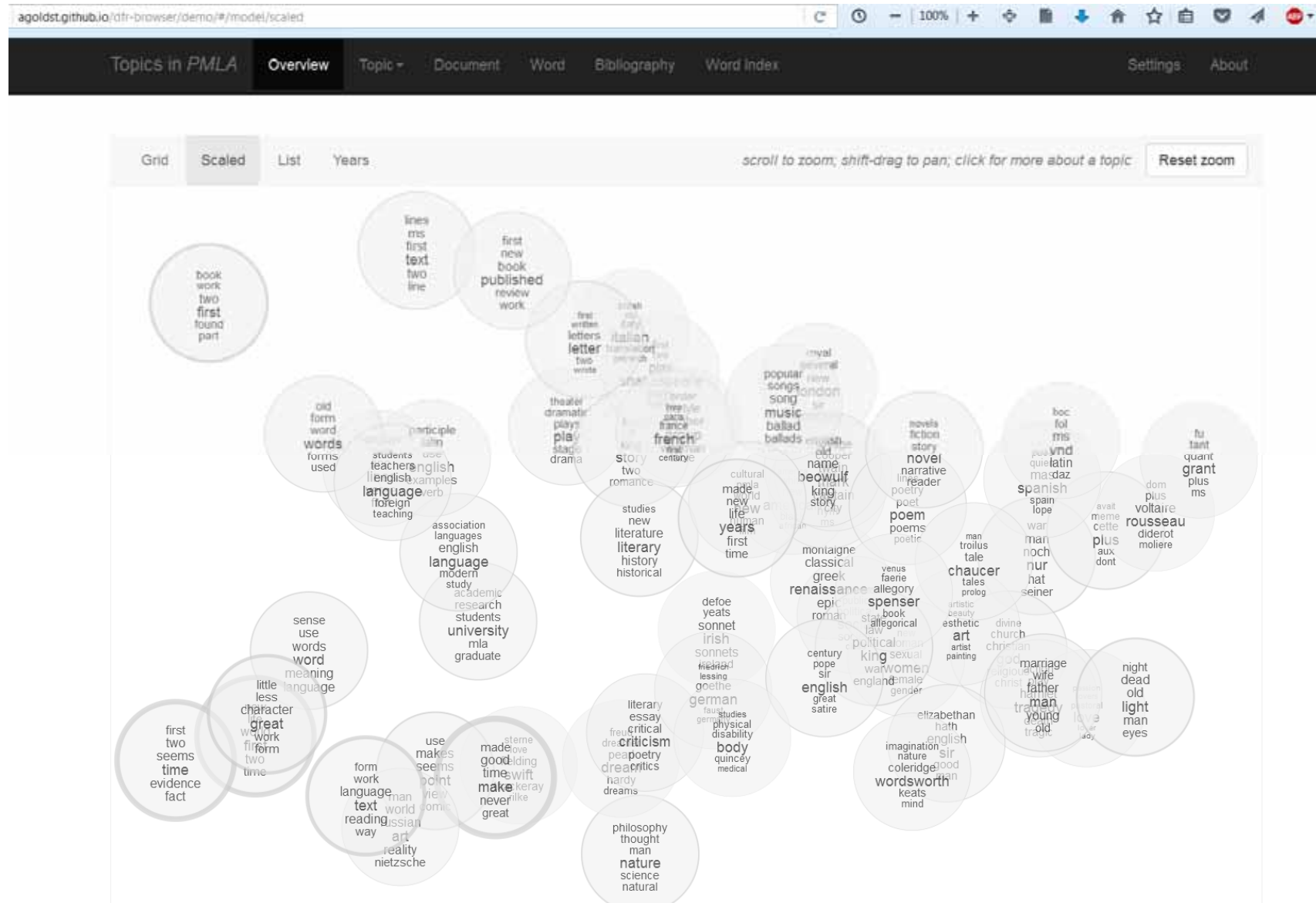
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

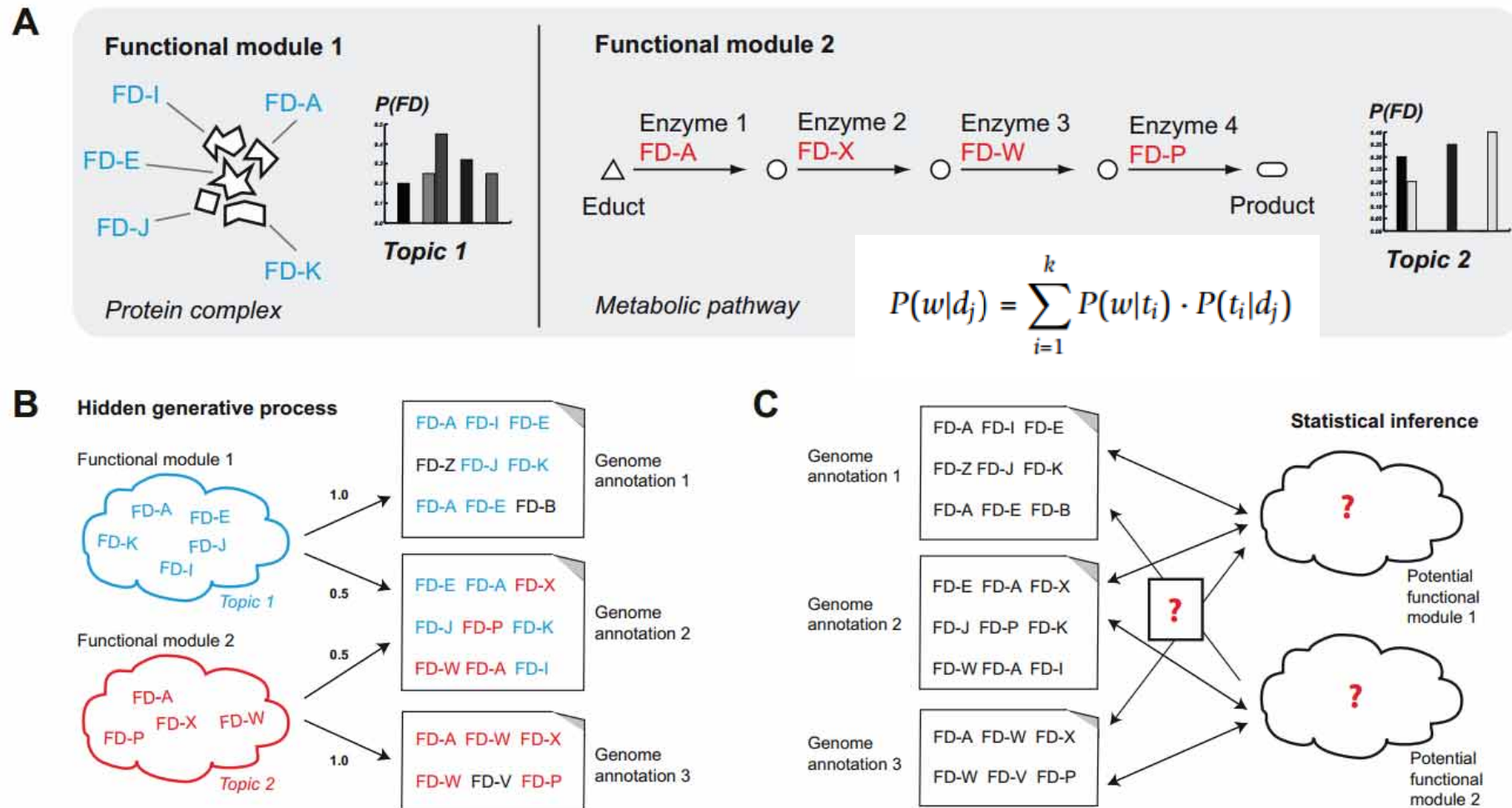
Given the parameters α and β , the joint distribution of a topic mixture θ , a set of N topics z , and a set of N words w is given by:

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

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

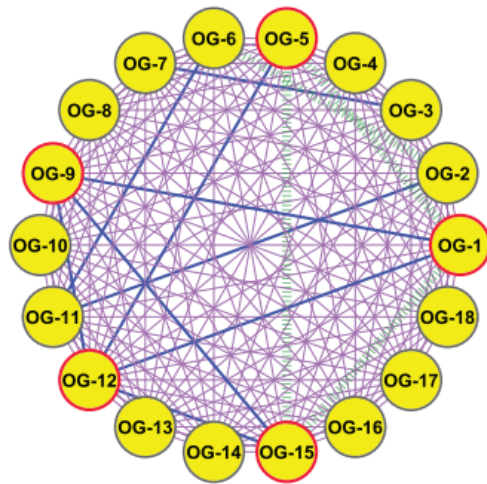


<http://agoldst.github.io/dfr-browser/demo/#/model/scaled>

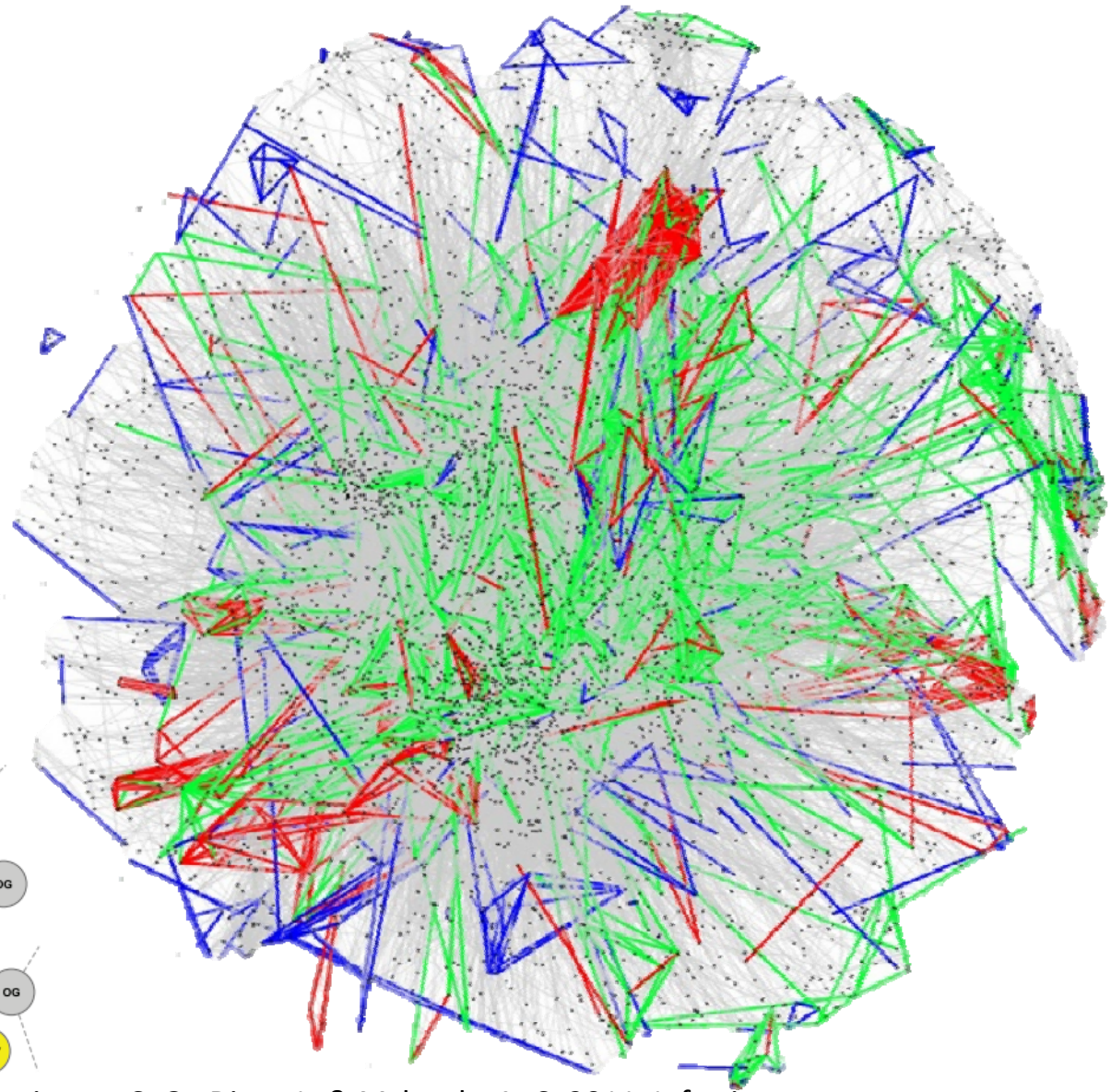
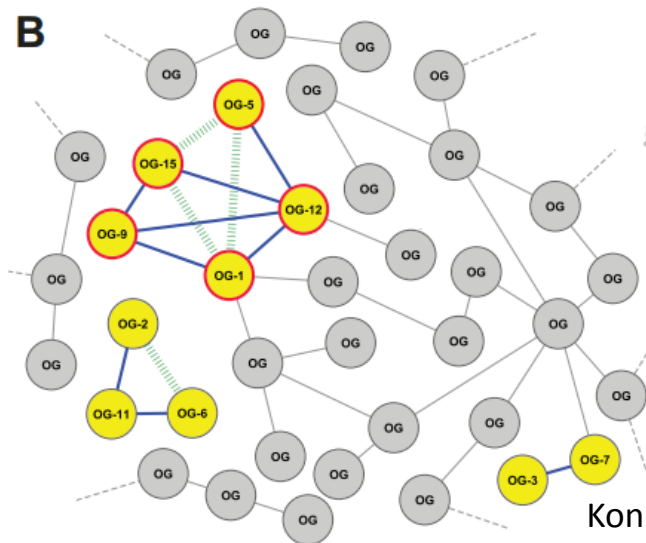


Konietzny, S. G., Dietz, L. & Mchardy, A. C. 2011. Inferring functional modules of protein families with probabilistic topic models. BMC bioinformatics, 12, (1), 1.

A



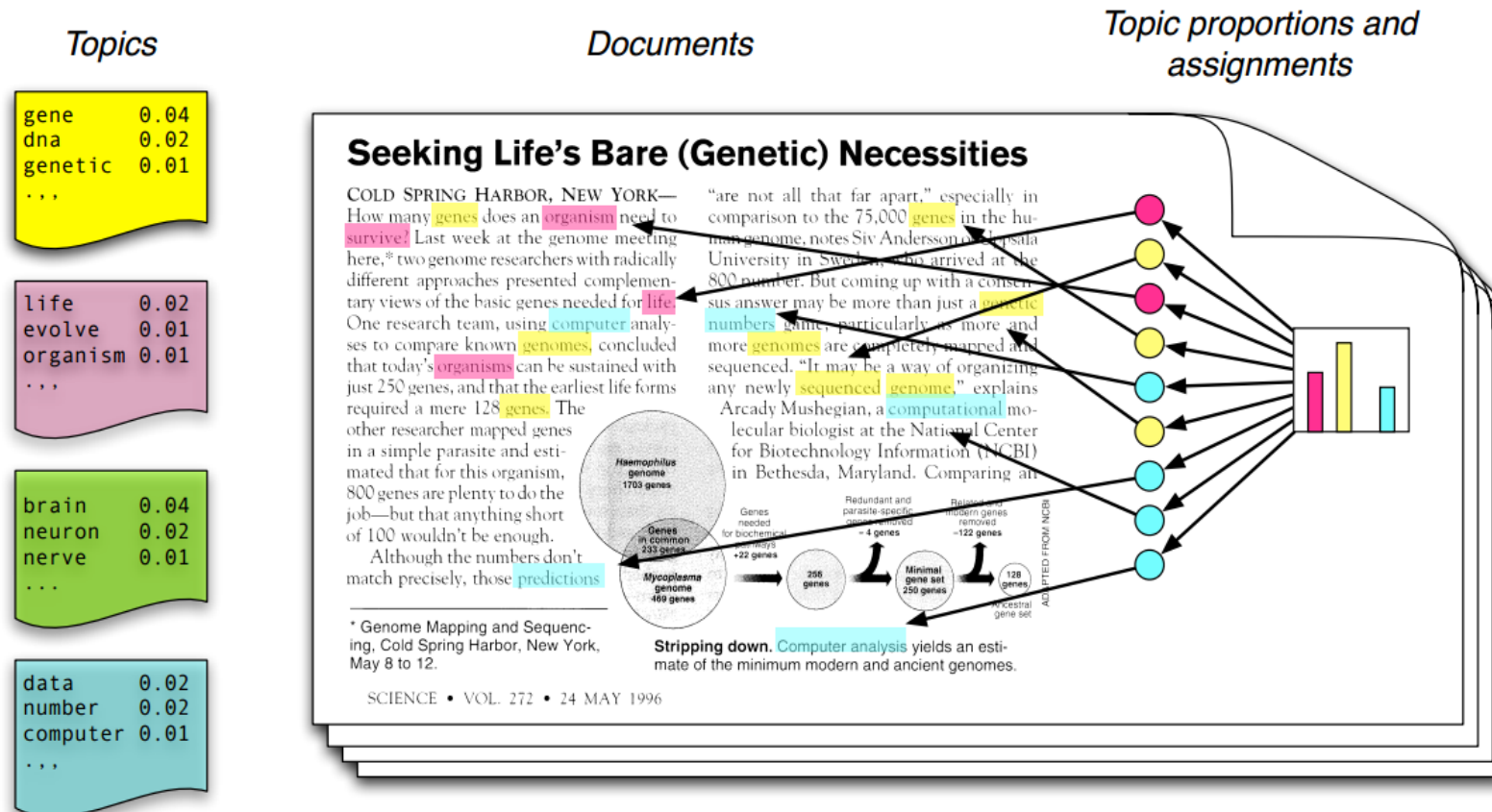
B



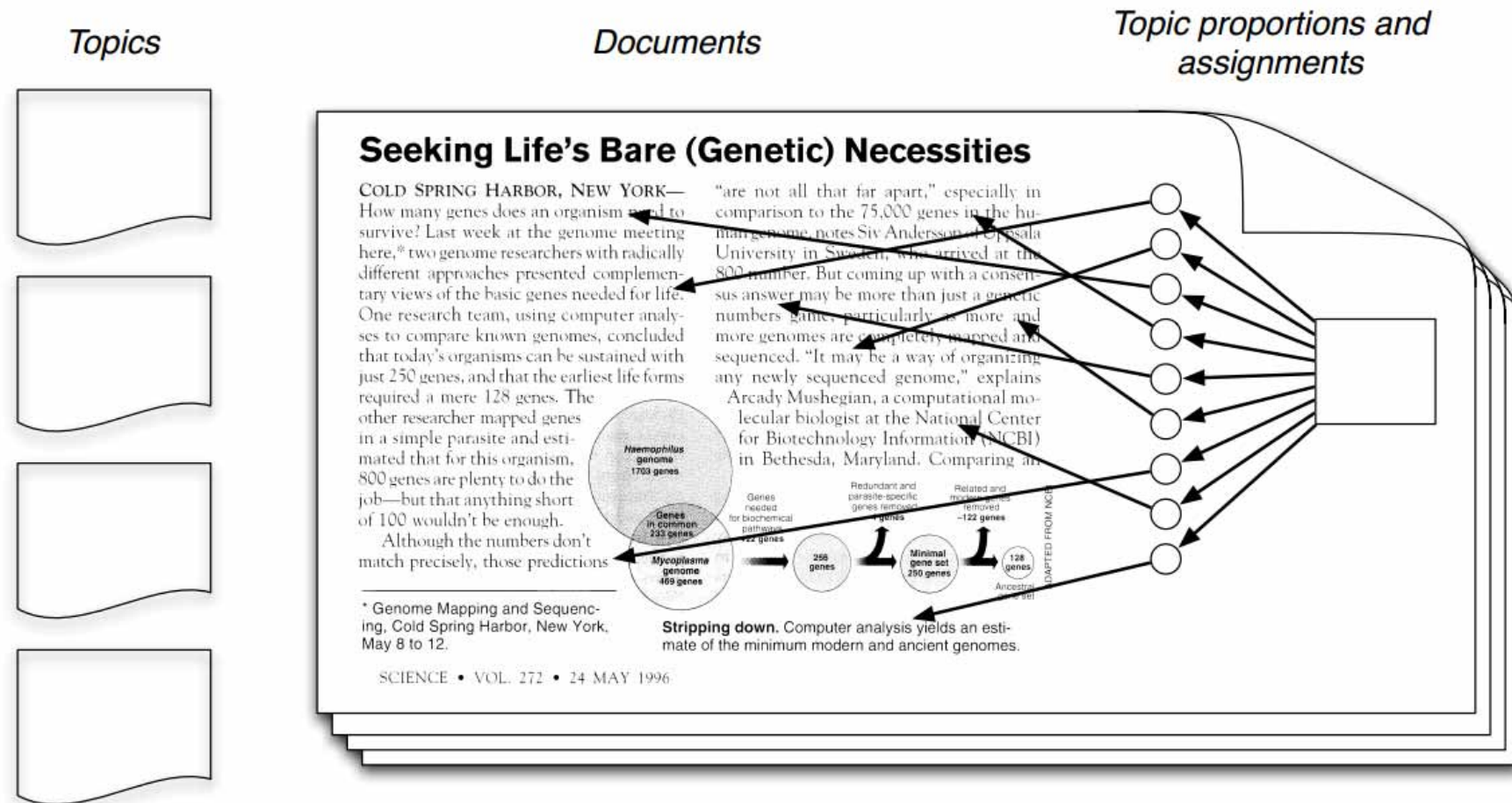
Konietzny, S. G., Dietz, L. & Mchardy, A. C. 2011. Inferring functional modules of protein families with probabilistic topic models. BMC bioinformatics, 12, (1), 1.

Goal: to get insight in unknown document collections

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

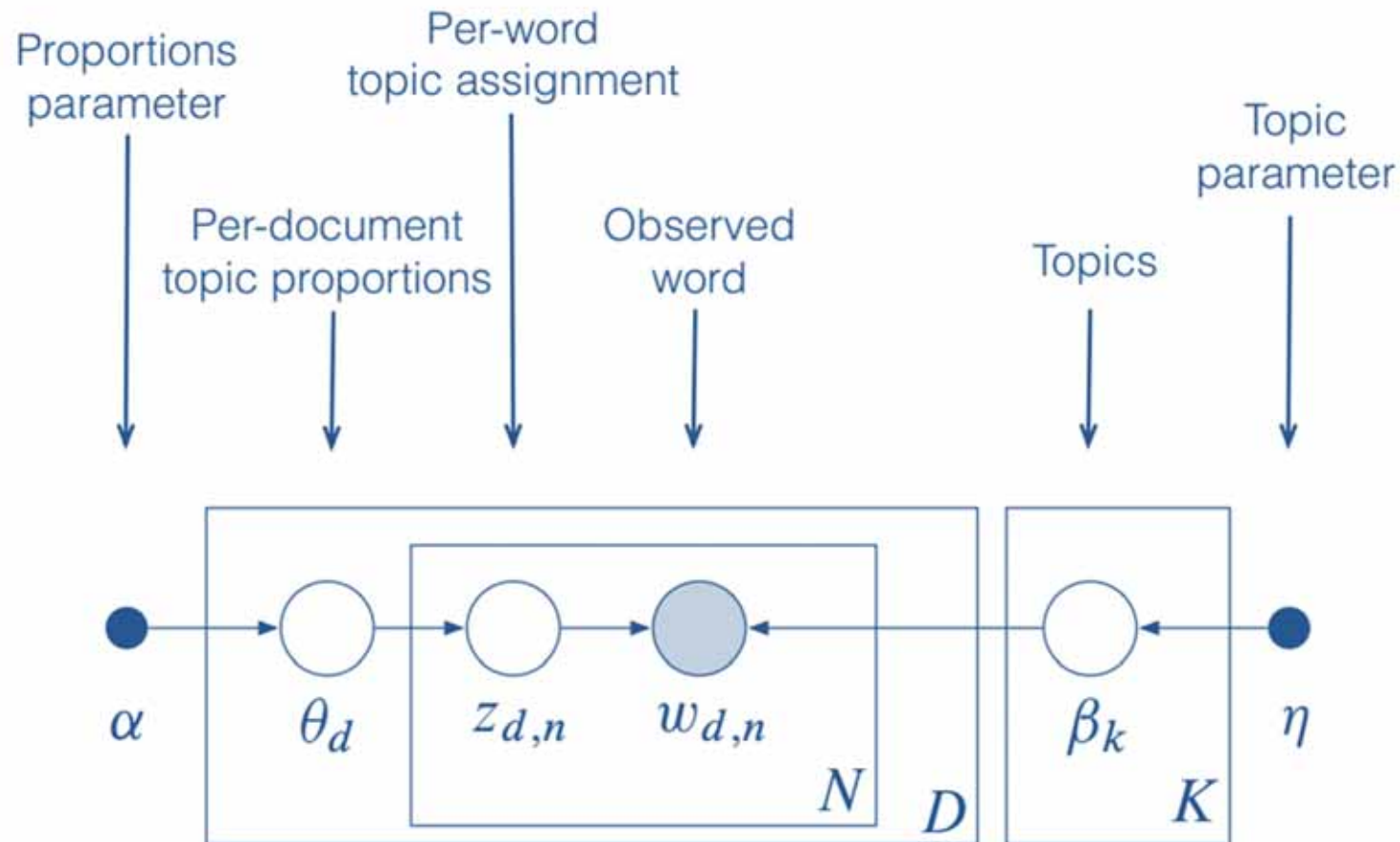


Each doc is a random mix of corpus-wide topics
and each word is drawn from one of these topics

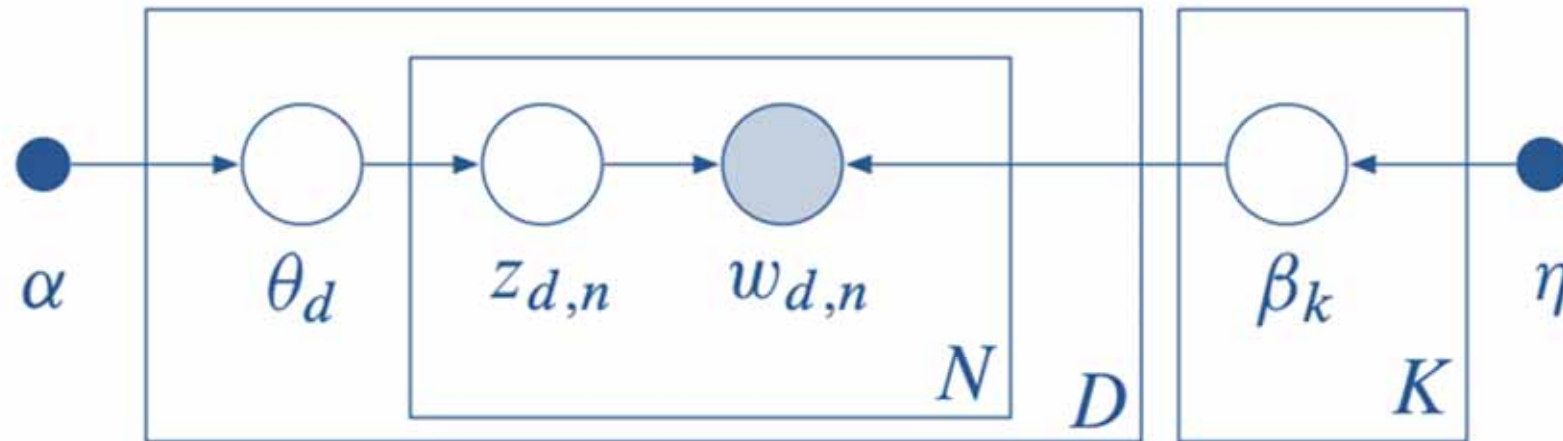


We only observe the docs – the other structure is hidden; then we compute the posterior $p(t,p,a | docs)$

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

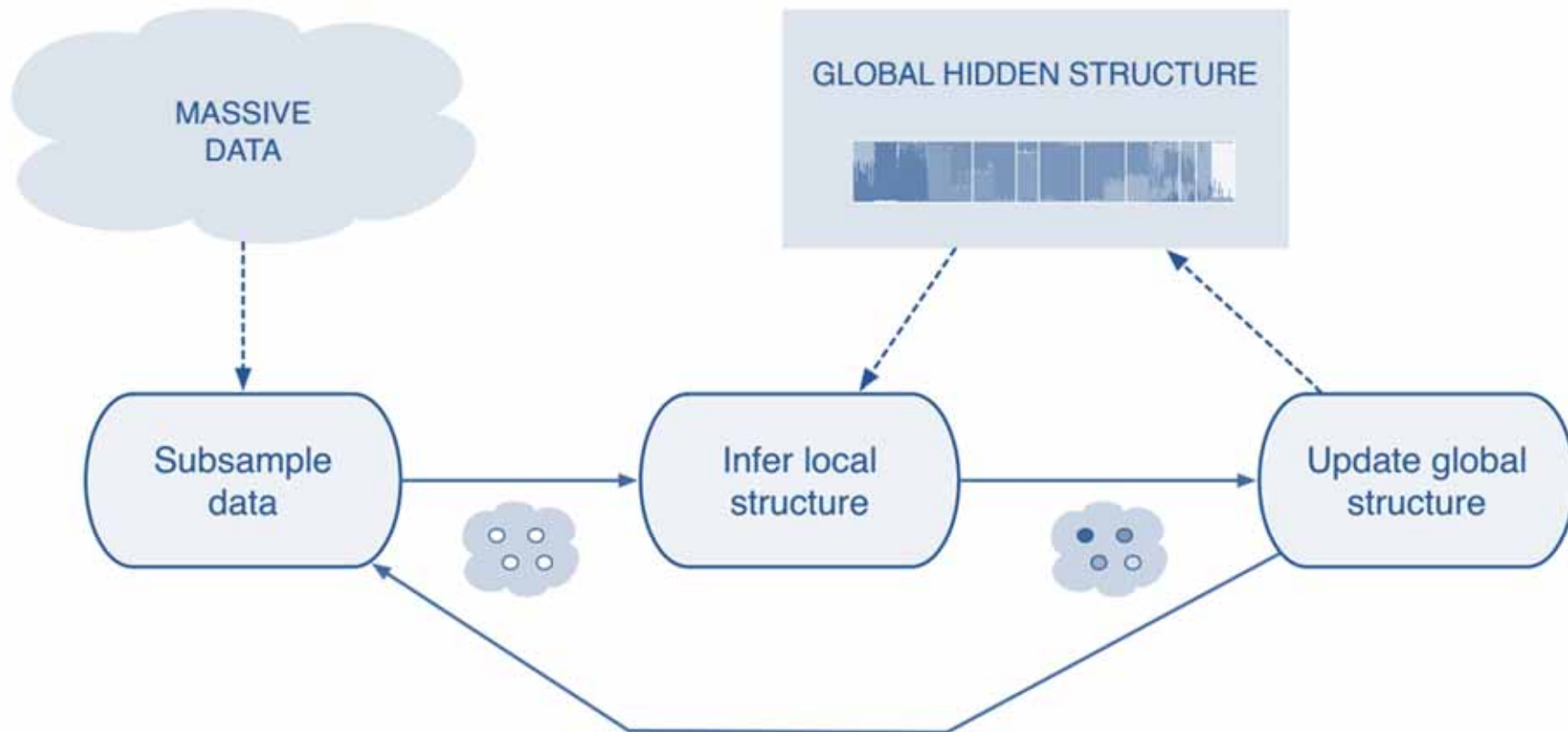


- Encodes assumptions on data with a factorization of the joint
- Connects assumptions to algorithms for computing with data
- Defines the posterior (through the joint)



$$p(\beta, \theta, z | w) = \frac{p(\beta, \theta, z, w)}{\int_{\beta} \int_{\theta} \sum_z p(\beta, \theta, z, w)}$$

We can't compute the denominator, the marginal $p(w)$, therefore we use **approximate inference**;
However, this do not scale well ...



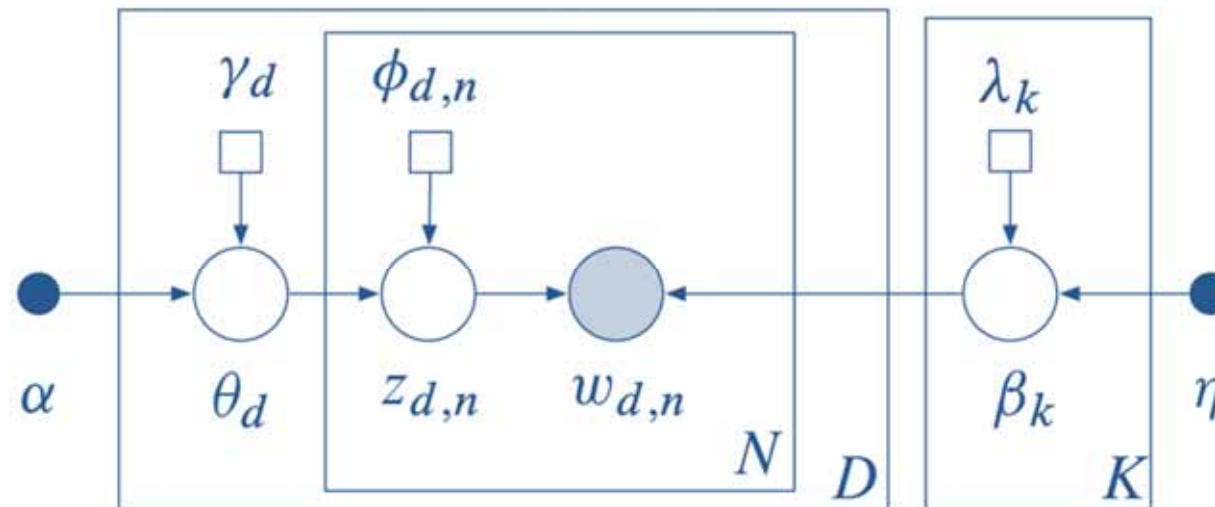
Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. 2013. Stochastic variational inference. The Journal of Machine Learning Research, 14, (1), 1303-1347.


```
1: Initialize  $\lambda^{(0)}$  randomly.
2: Set the step-size schedule  $\rho_t$  appropriately.
3: repeat
4:   Sample a document  $w_d$  uniformly from the data set.
5:   Initialize  $\gamma_{dk} = 1$ , for  $k \in \{1, \dots, K\}$ .
6:   repeat
7:     For  $n \in \{1, \dots, N\}$  set
        
$$\phi_{dn}^k \propto \exp \{ \mathbb{E}[\log \theta_{dk}] + \mathbb{E}[\log \beta_{k,w_{dn}}] \}, k \in \{1, \dots, K\}.$$

8:     Set  $\gamma_d = \alpha + \sum_n \phi_{dn}$ .
9:   until local parameters  $\phi_{dn}$  and  $\gamma_d$  converge.
10:  For  $k \in \{1, \dots, K\}$  set intermediate topics
        
$$\hat{\lambda}_k = \eta + D \sum_{n=1}^N \phi_{dn}^k w_{dn}.$$

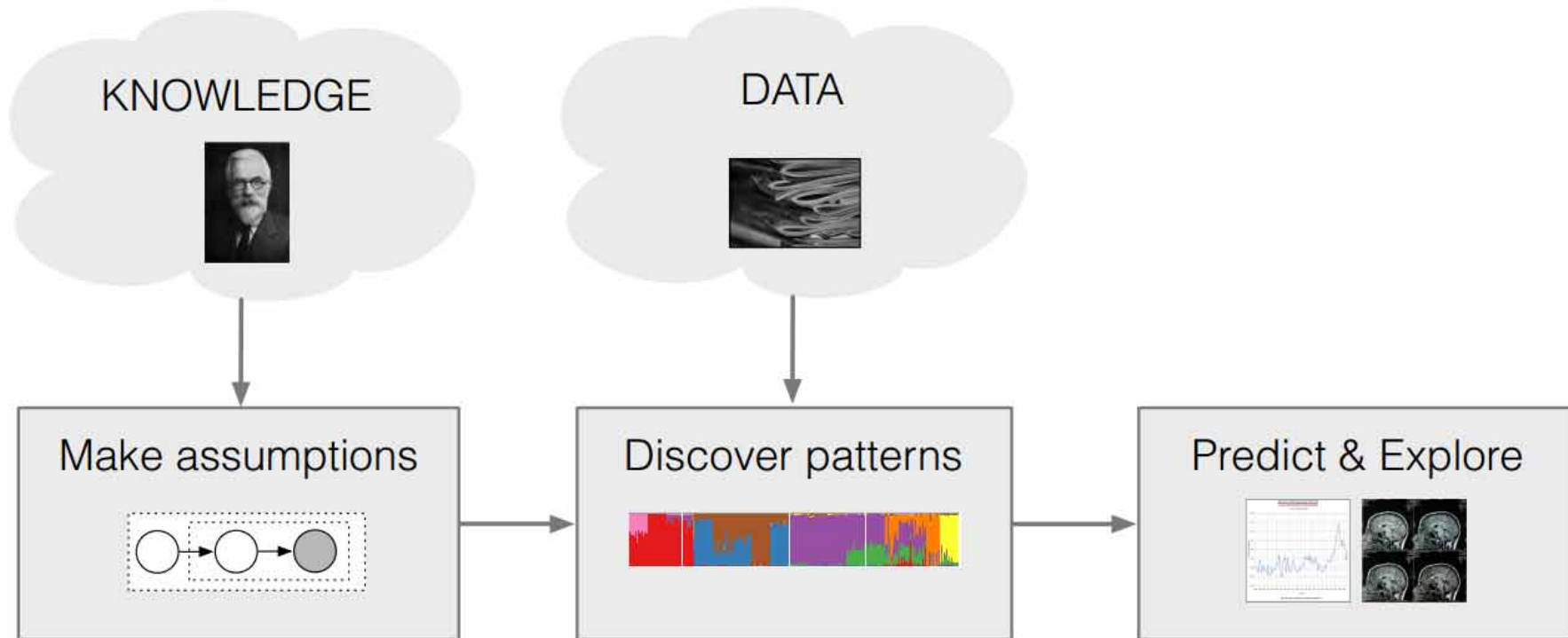
11:  Set  $\lambda^{(t)} = (1 - \rho_t) \lambda^{(t-1)} + \rho_t \hat{\lambda}$ .
12: until forever
```

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

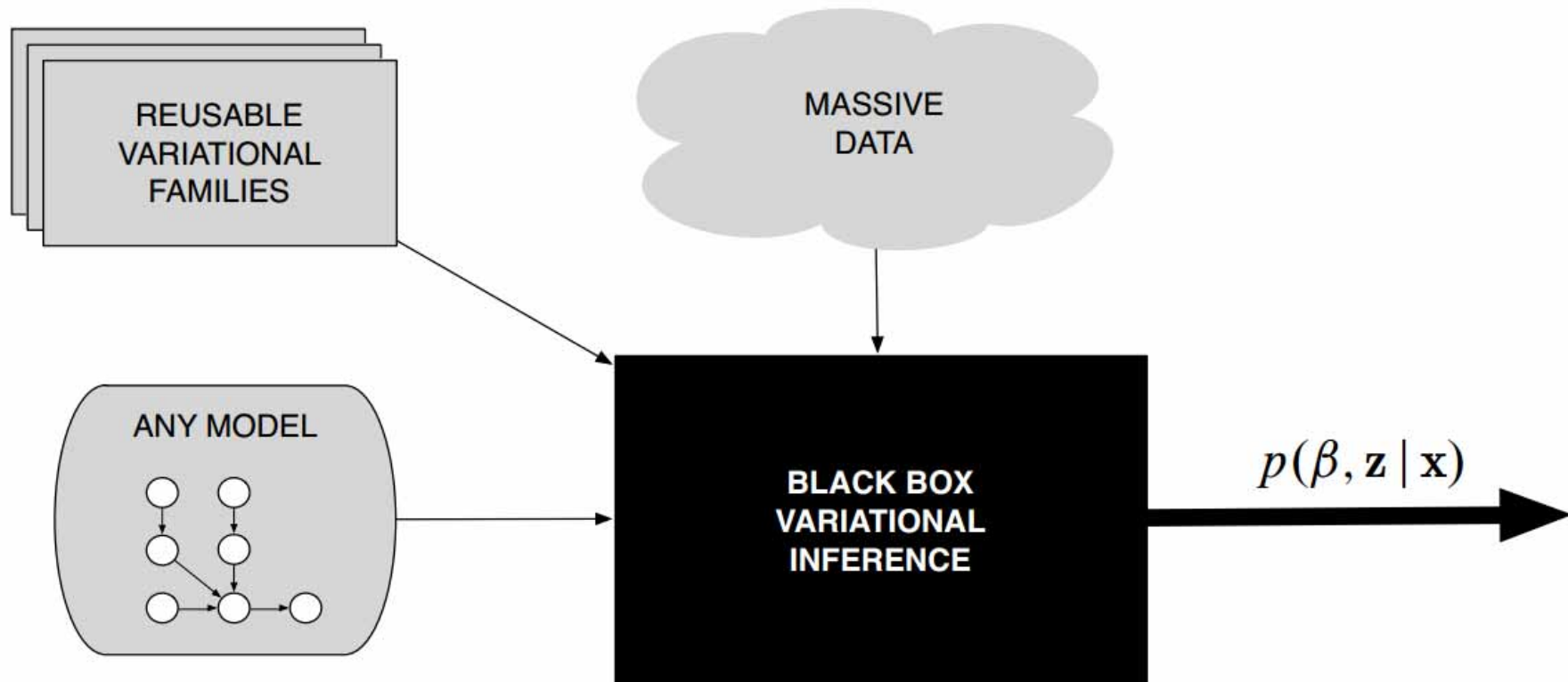


1. Sample a document
2. Estimate the local variational parameters using the current topics
3. Form intermediate topics from those local parameters
4. Update topics as a weighted average of intermediate and current topics

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



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



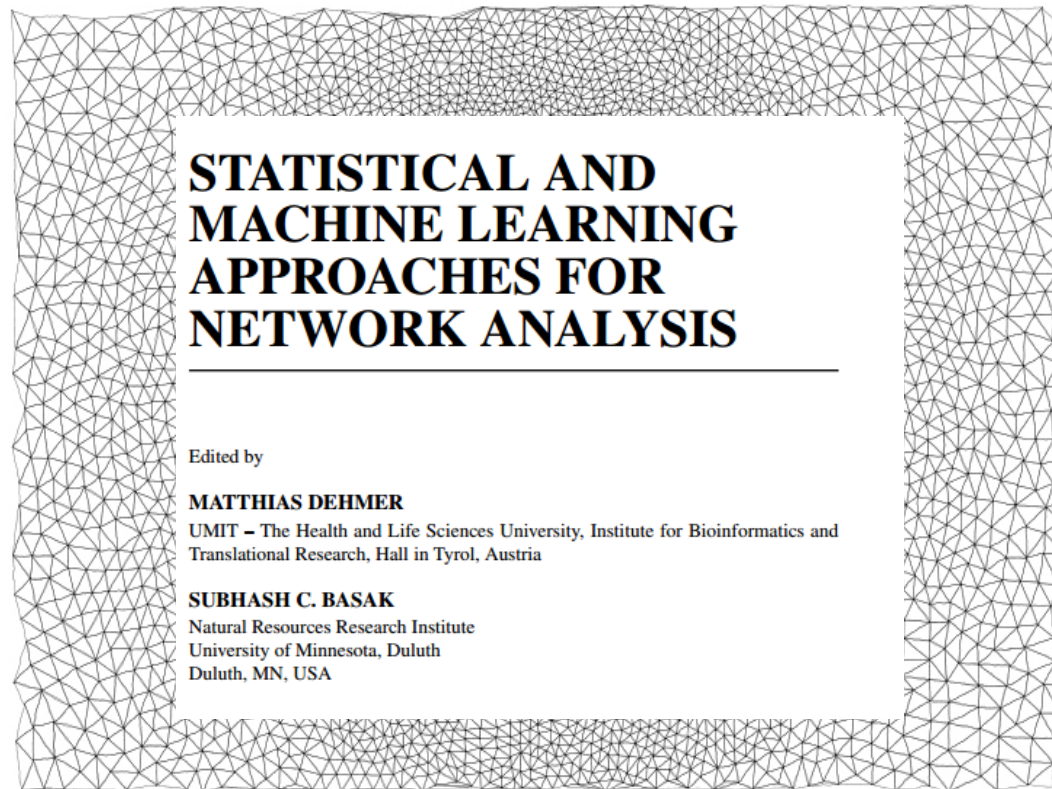
Hoffman, M. D., Blei, D. M., Wang, C. & Paisley, J. 2013. Stochastic variational inference. The Journal of Machine Learning Research, 14, (1), 1303-1347.

- Flexible and expressive components for building models
- Scalable and generic inference algorithms
- Easy to use software to stretch probabilistic modeling into the health domain
- Topic models are only one approach towards detection of topics in text collections
- More general: Identify re-occurring patterns in data collections generally ...
- Much open work for you in the future 😊

- 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/>

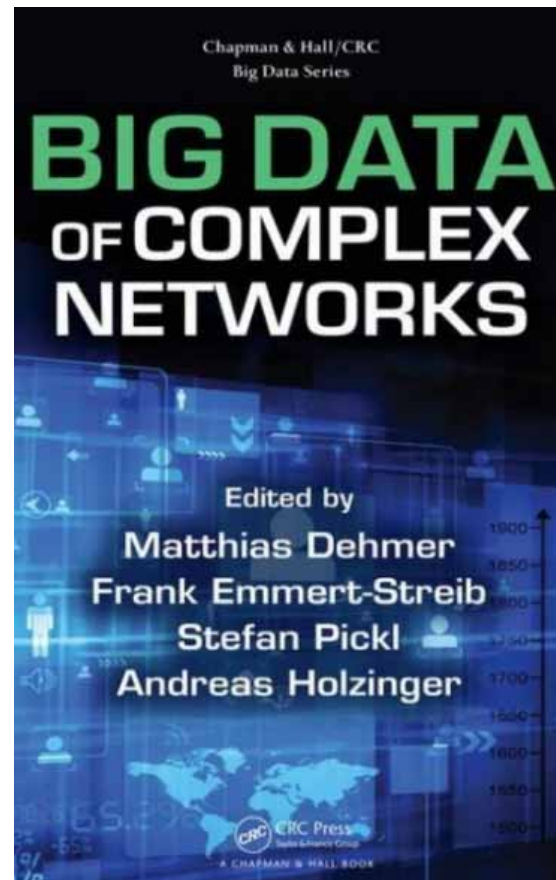
03 Knowledge Representation in Network Medicine

Networks = Graphs



<http://www.wired.com/tag/network-science/>

http://www.barabasilab.com/pubs/CCNR-ALB_Publications/200907-24_Science-Decade/200907-24_Science-CoverImage.gif



Dehmer, M., Emmert-Streib, F., Pickl, S. & Holzinger, A. (eds.) 2016. Big Data of Complex Networks, Boca Raton, London, New York: CRC Press Taylor & Francis Group.

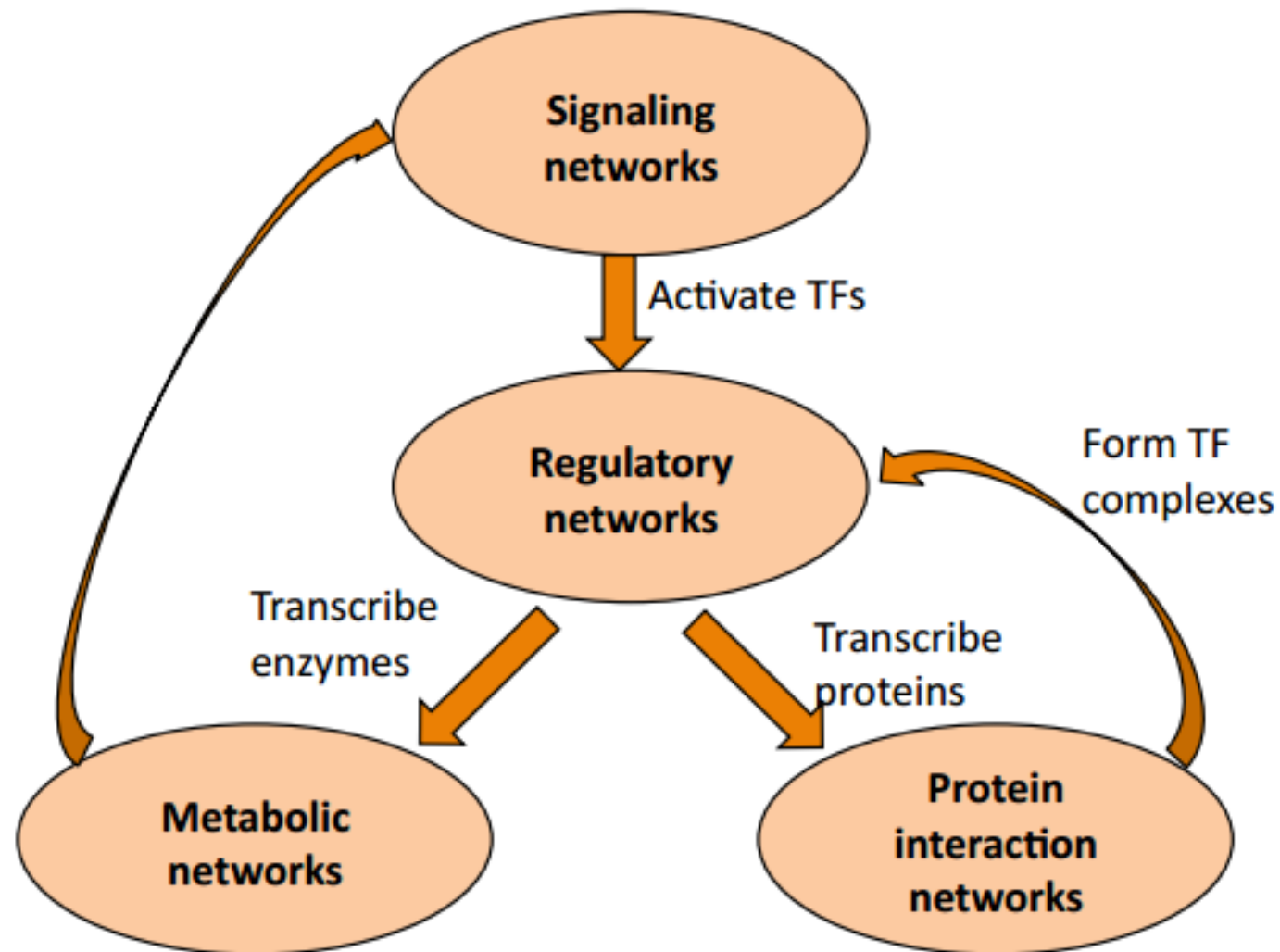


Image credit to Anna Goldenberg, Toronto

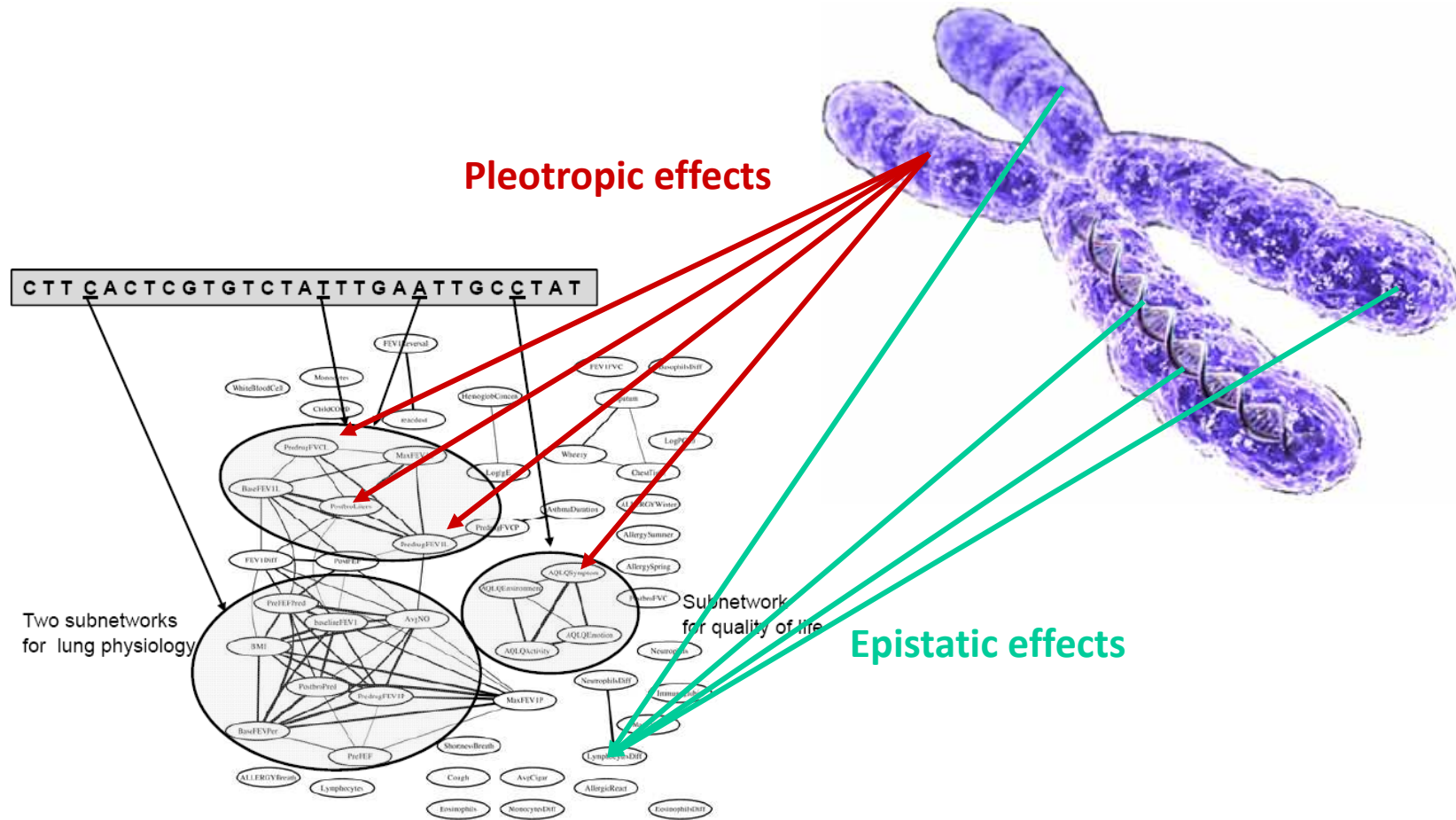
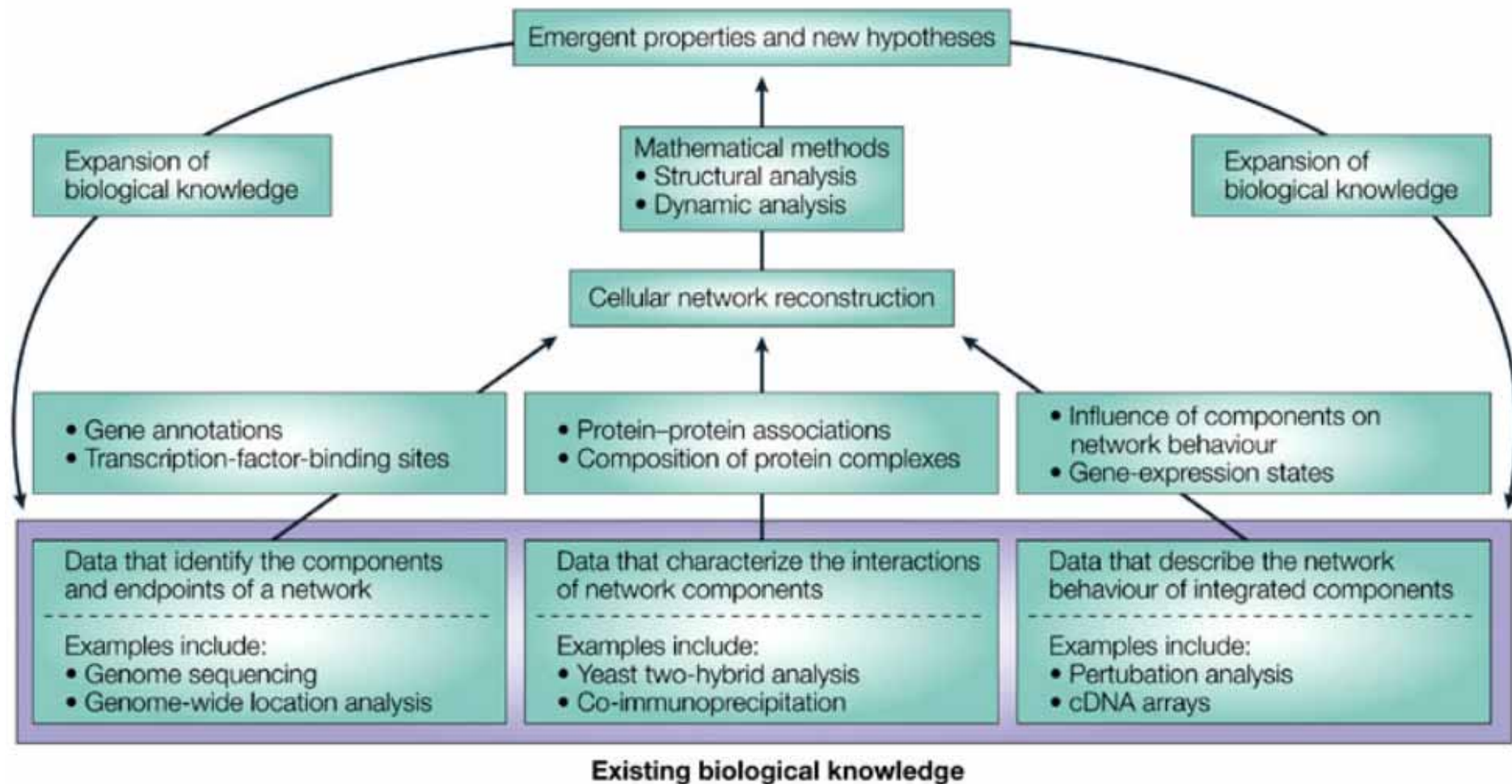


Image credit to Eric Xing, Carnegie Mellon University, Pittsburgh



Nature Reviews | Molecular Cell Biology

Image description find here:

http://www.nature.com/nrm/journal/v6/n2/fig_tab/nrm1570_F1.html

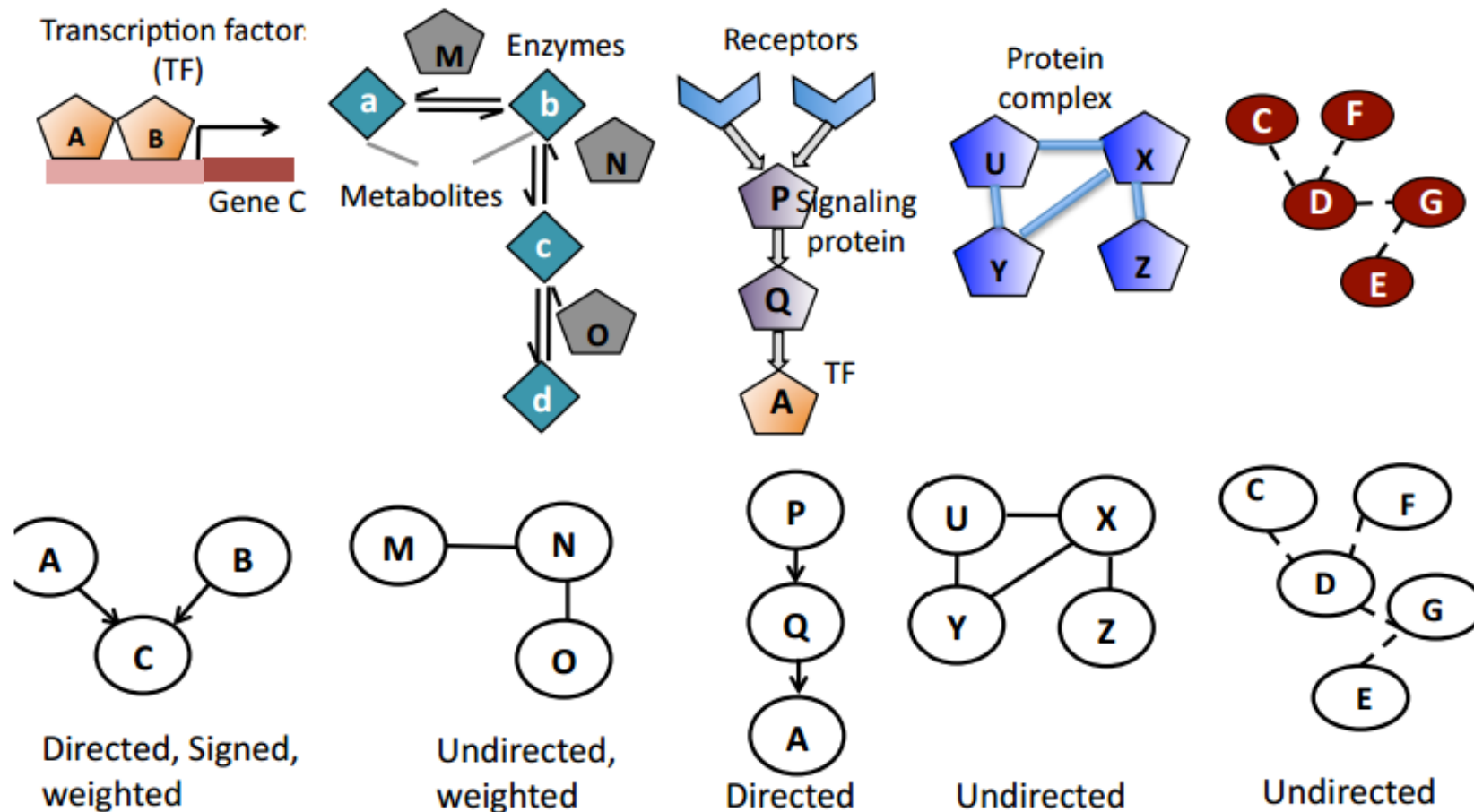
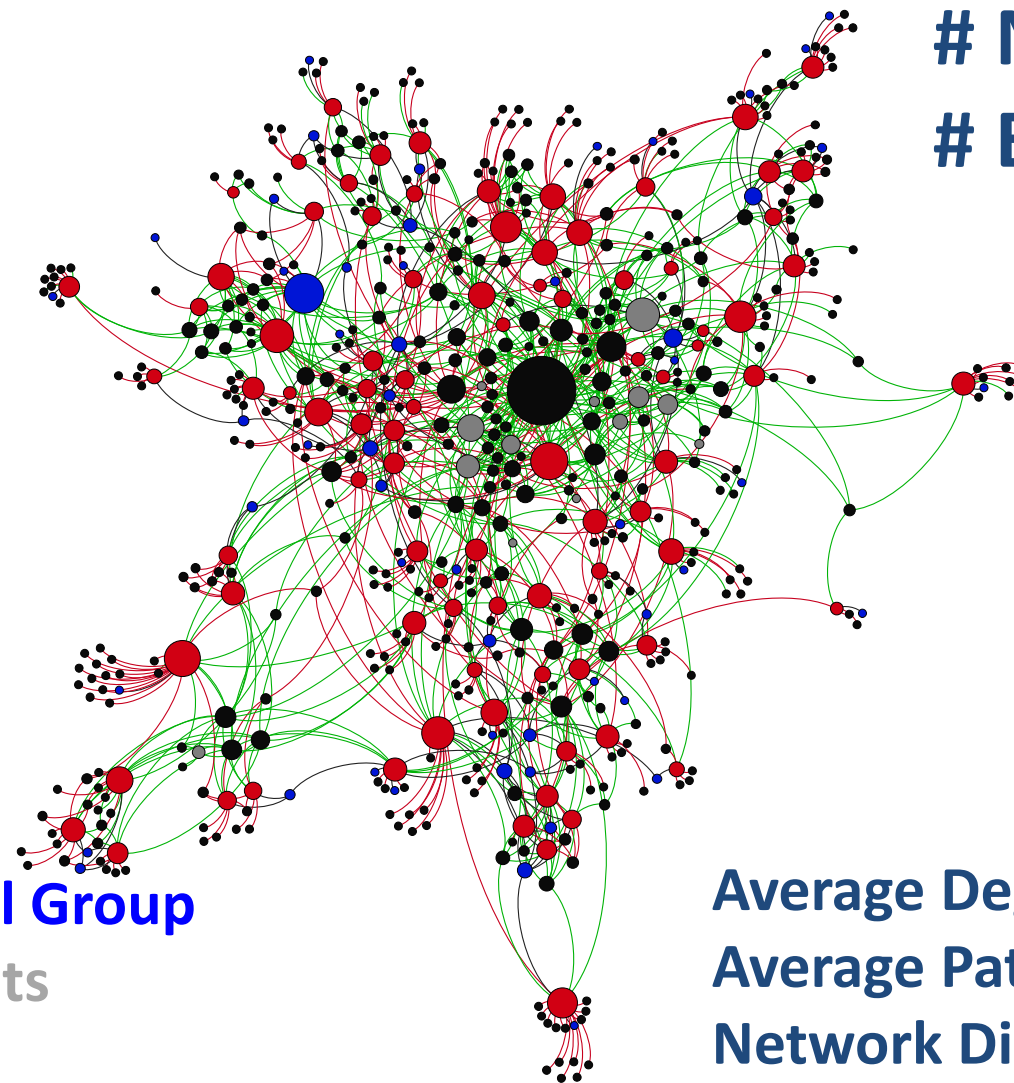


Image credit to Anna Goldenberg, Toronto



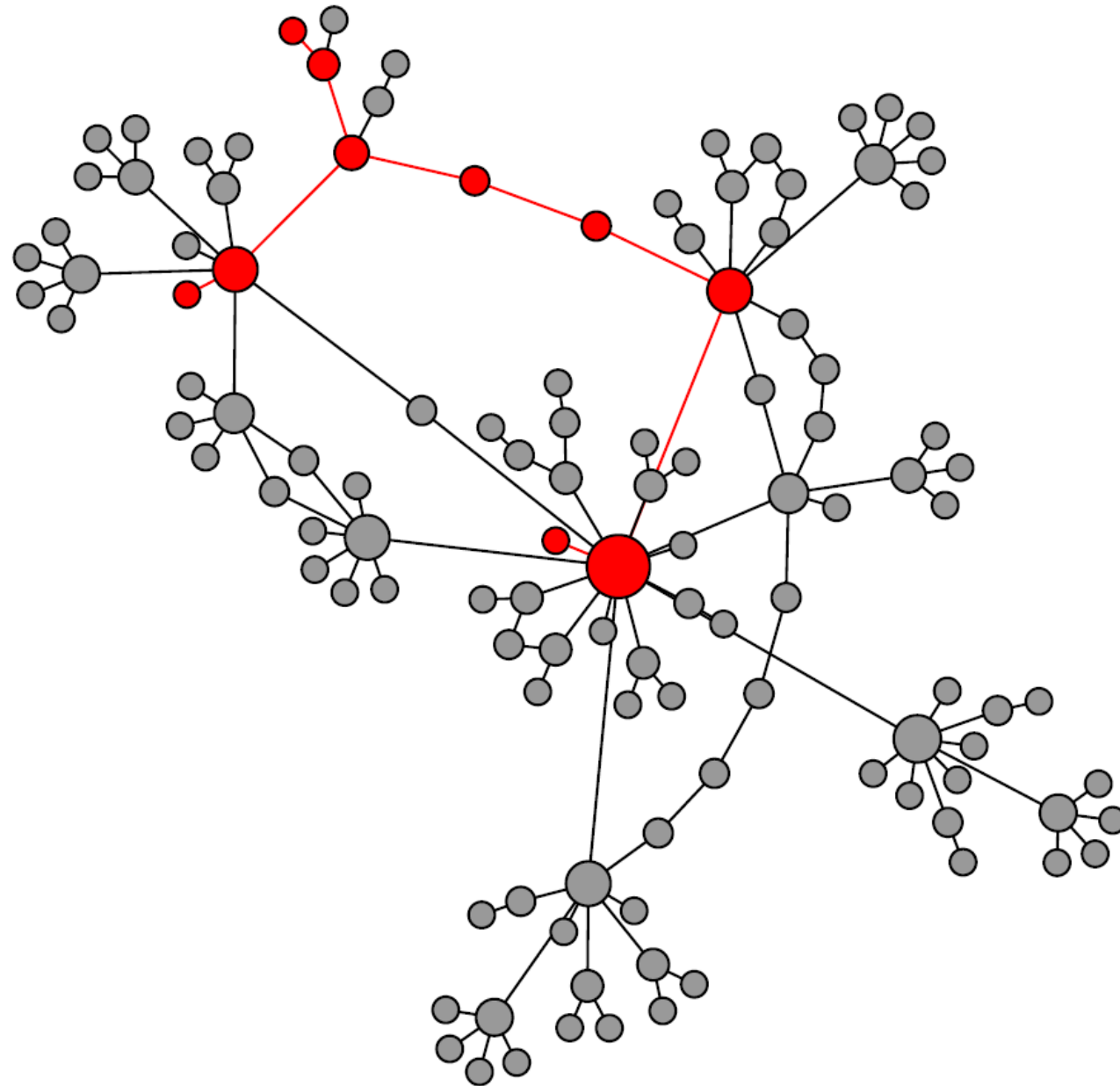
Nodes: 641
Edges: 1250

Agent
Condition
Pharmacological Group
Other Documents

Average Degree: 3.888
Average Path Length: 4.683
Network Diameter: 9

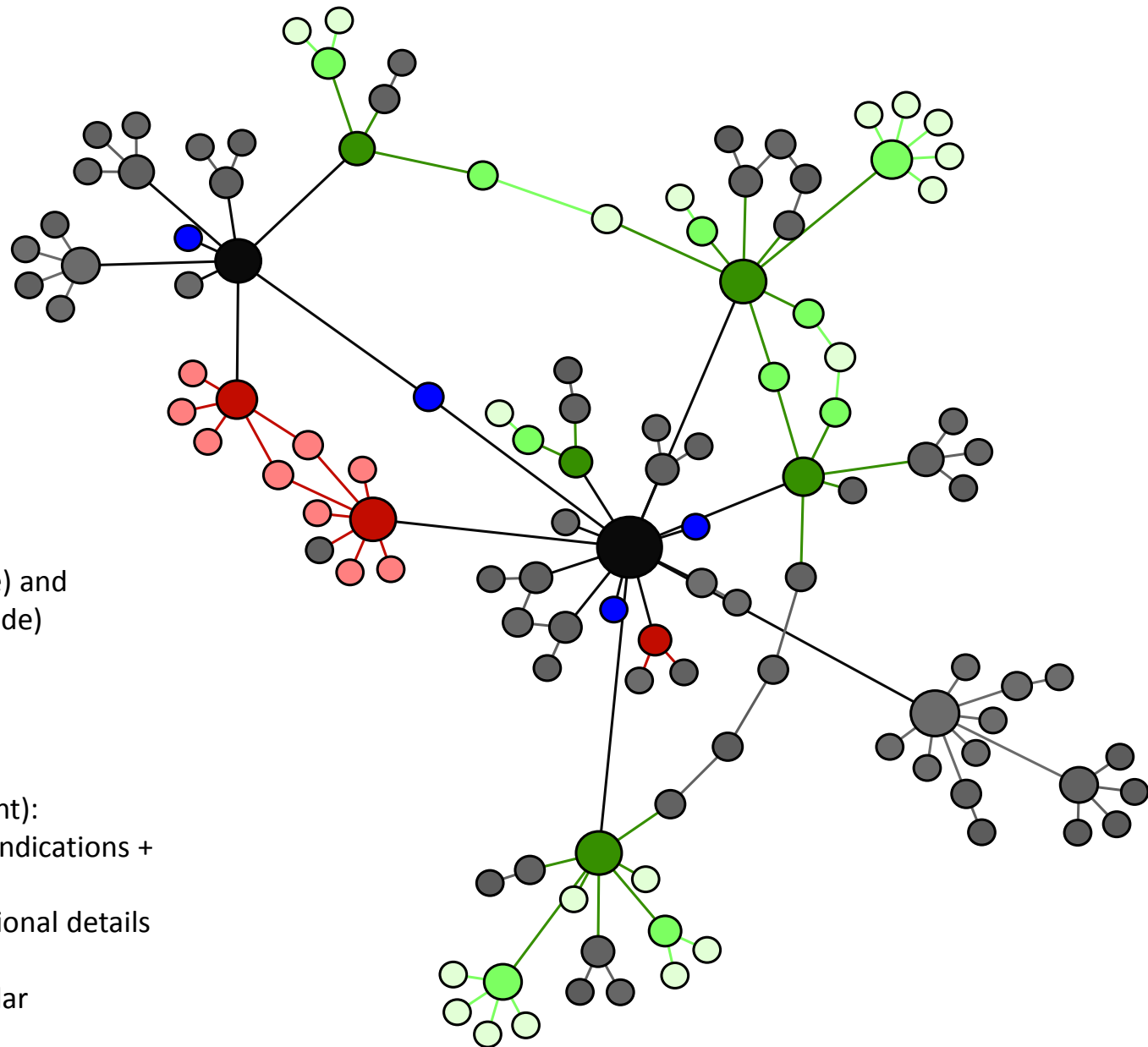
Holzinger, A., Ofner, B., Dehmer, M.: Multi-touch Graph-Based Interaction for Knowledge Discovery on Mobile Devices: State-of-the-Art and Future Challenges. In: LNCS 8401, pp. 241–254, (2014)

- **Nodes**
 - drugs
 - clinical guidelines
 - **patient conditions** (indication, contraindication)
 - **pharmacological groups**
 - tables and calculations of medical scores
 - algorithms and other medical documents
- **Edges:** 3 crucial types of relations inducing medical relevance between two active substances
 - pharmacological groups
 - indications
 - contra-indications

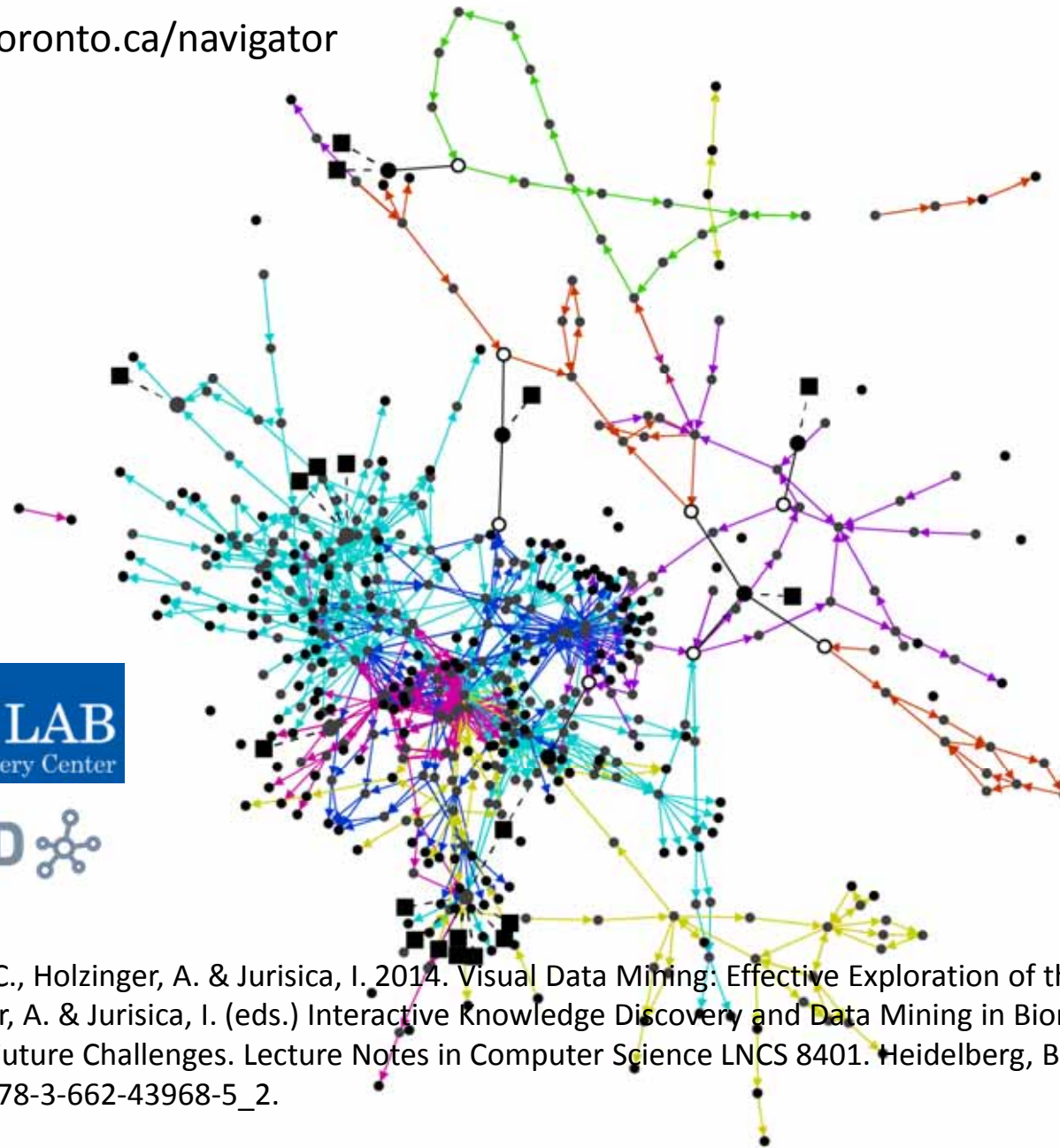


Relationship between
Adrenaline (center black node) and
Dobutamine (top left black node)
Blue: Pharmacological Group
Dark red: Contraindication;
Light red: Condition

Green nodes (from dark to light):
1. Application (one ore more indications +
corresponding dosages)
2. Single indication with additional details
(e. g. "VF after 3rd Shock")
3. Condition (e.g. VF, Ventricular
Fibrillation)



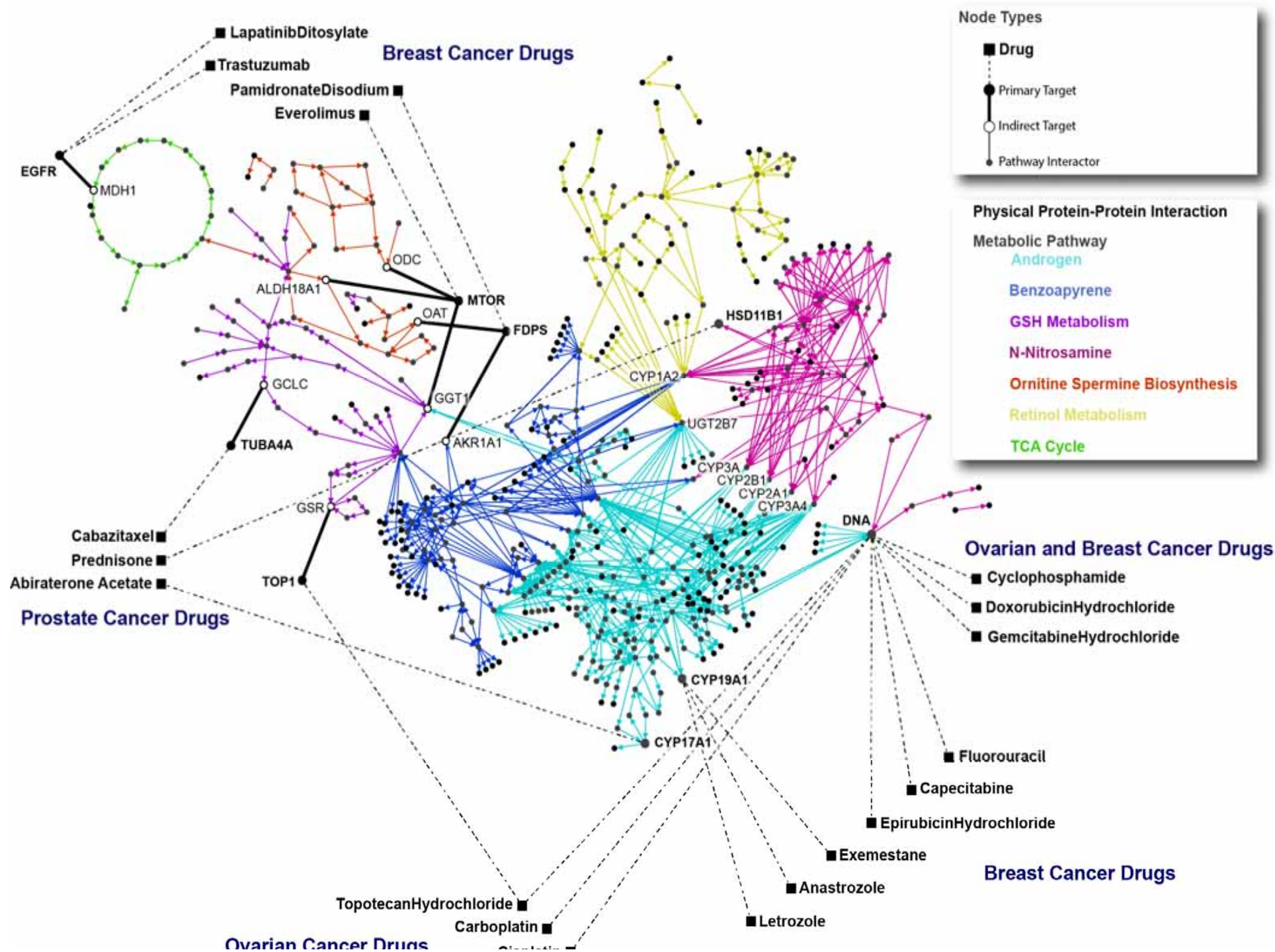
<http://ophid.utoronto.ca/navigator>

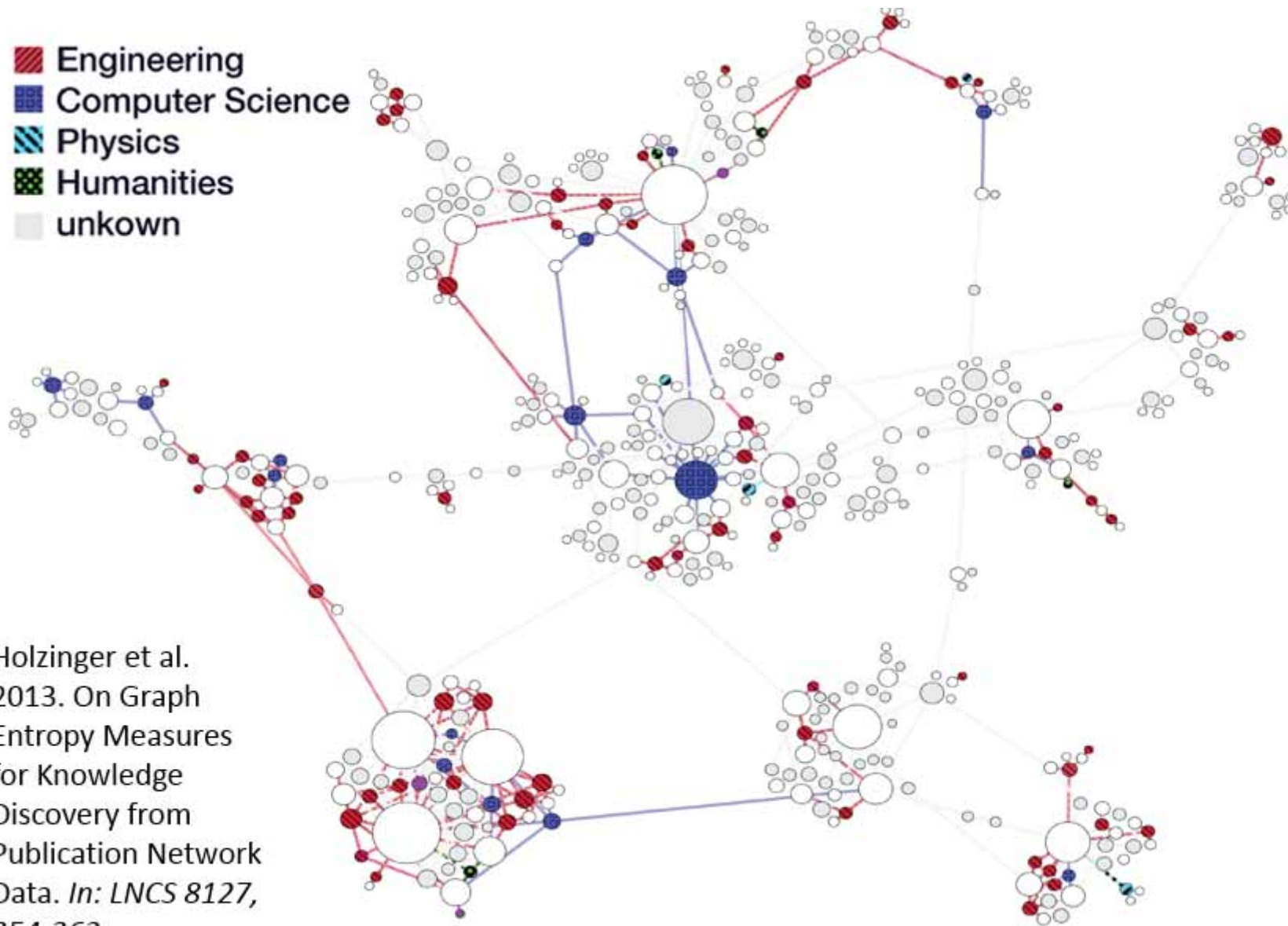


JURISICA LAB
IBM Life Sciences Discovery Center



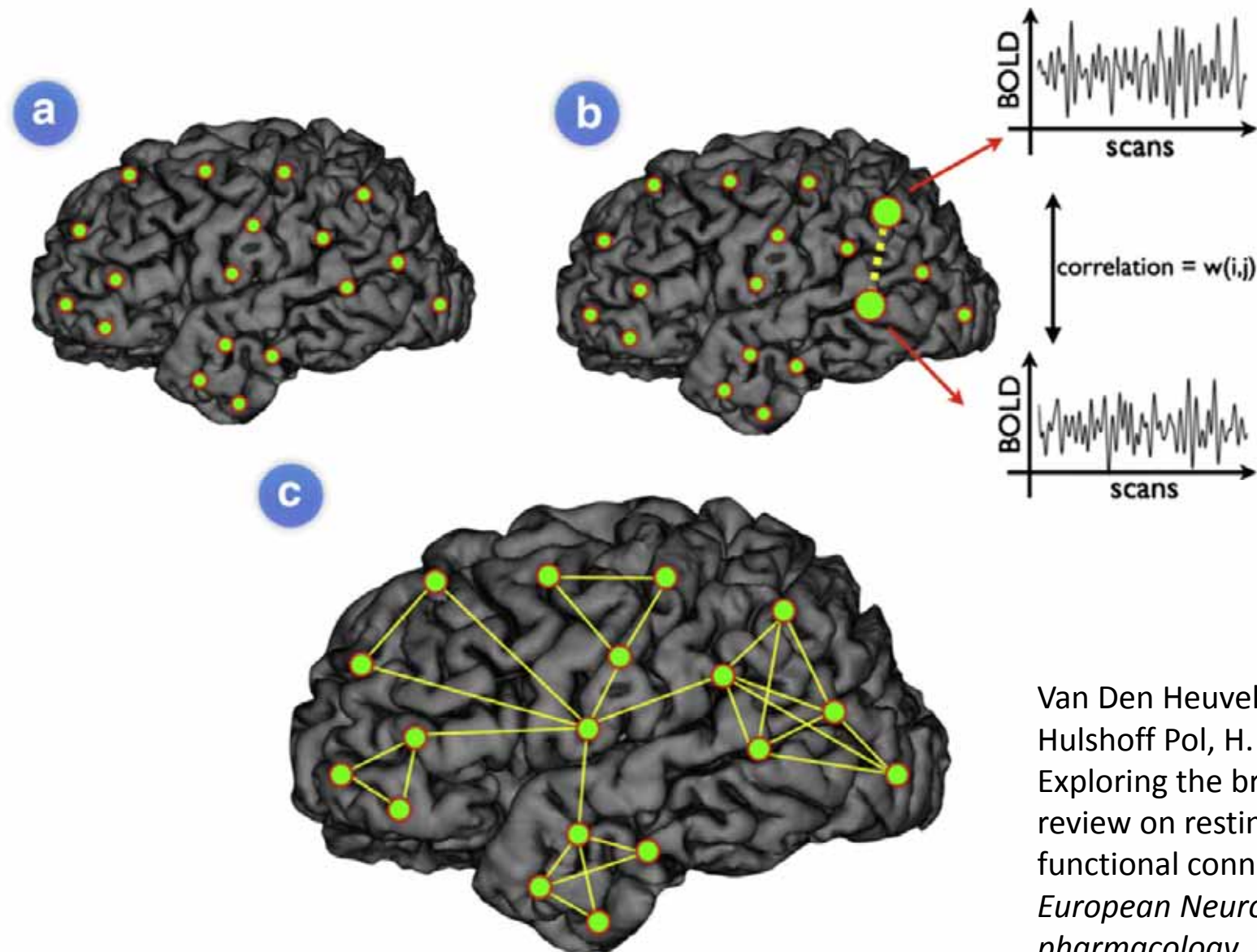
Otasek, D., Pastrello, C., Holzinger, A. & Jurisica, I. 2014. Visual Data Mining: Effective Exploration of the Biological Universe. In: Holzinger, A. & Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-of-the-Art and Future Challenges. Lecture Notes in Computer Science LNCS 8401. Heidelberg, Berlin: Springer, pp. 19–34, doi:10.1007/978-3-662-43968-5_2.





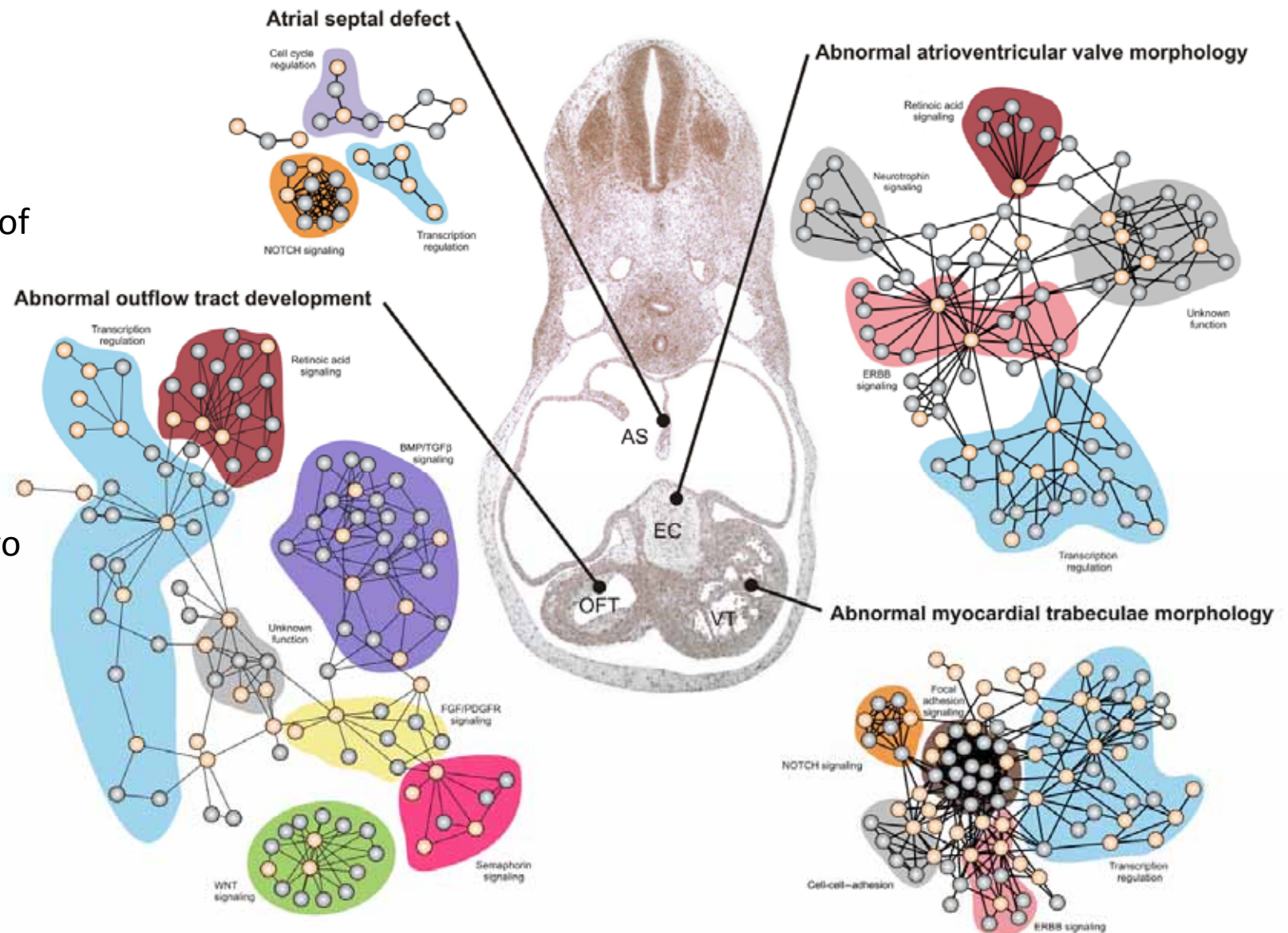
- **Problem:** What is the max. number of edges of an Relative Neighborhood Graph in R^3 ? No supra-linear lower bound is known.
- **Problem:** What is the structural interpretation of graph measures ? They are mappings which maps graphs to the reals. Thus, they can be understood as graph complexity measures and investigating their structural interpretation relates to understand what kind of structural complexity they detect.
- **Problem:** It is important to visualize large networks meaningfully. So far, there has been a lack of interest to develop efficient software beyond the available commercial software.
- **Problem:** Are multi-touch interaction graphs structurally similar to other graphs (from known graph classes)? This calls for a comparison of graph classes and their structural characteristics.
- **Problem:** Which graph measures are suitable to determine the complexity of multi-touch interaction graphs? Does this lead to any meaningful classification based on their topology?
- **Problem:** What is interesting? Where to start the interaction?

Holzinger, A., Ofner, B., & Dehmer, M. (2014). Multi-touch Graph-Based Interaction for Knowledge Discovery on Mobile Devices: State-of-the-Art and Future Challenges. LNCS 8401 (pp. 241–254). Berlin, Heidelberg: Springer.

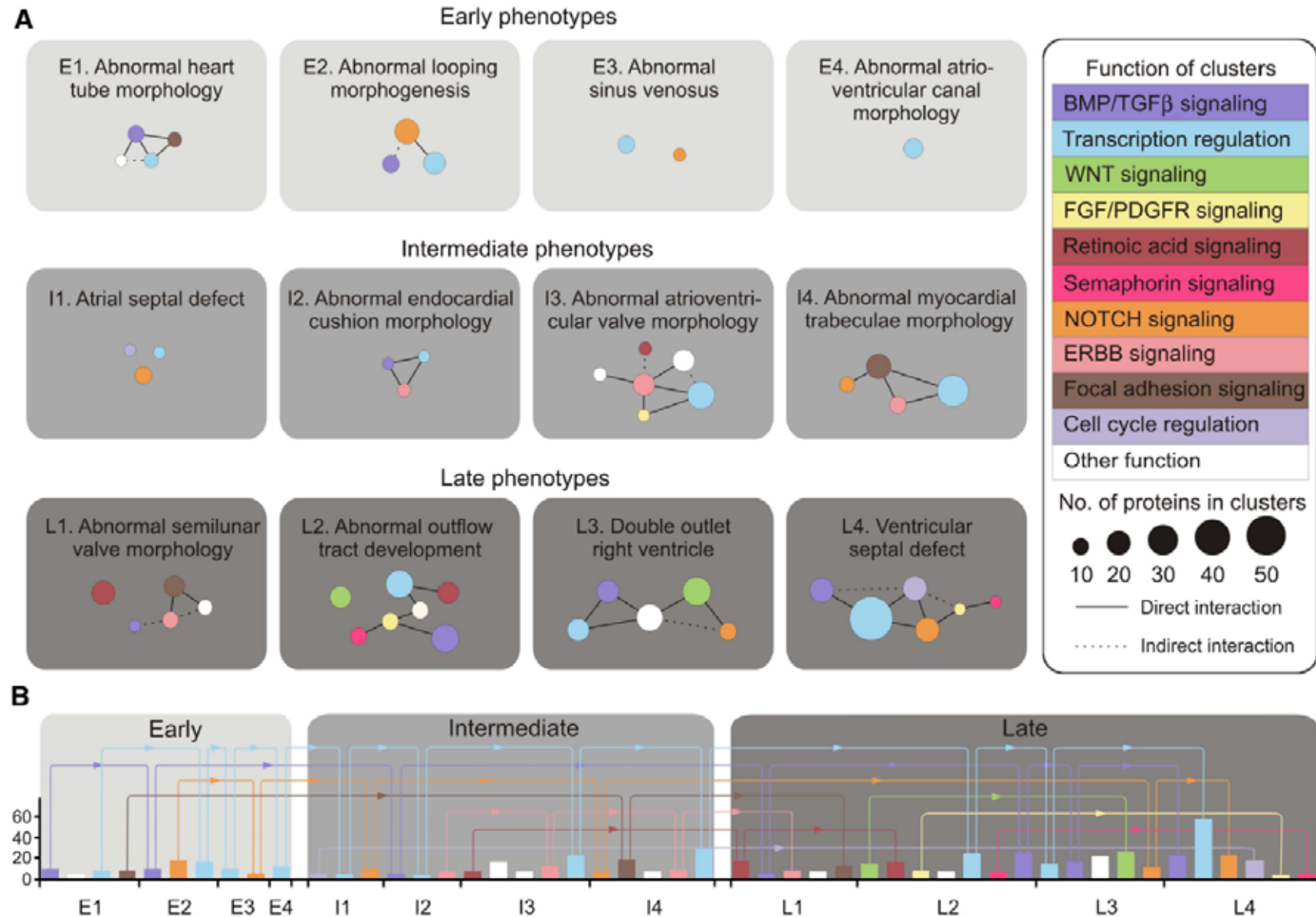


Van Den Heuvel, M. P. & Hulshoff Pol, H. E. (2010) Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20, 8, 519-534.

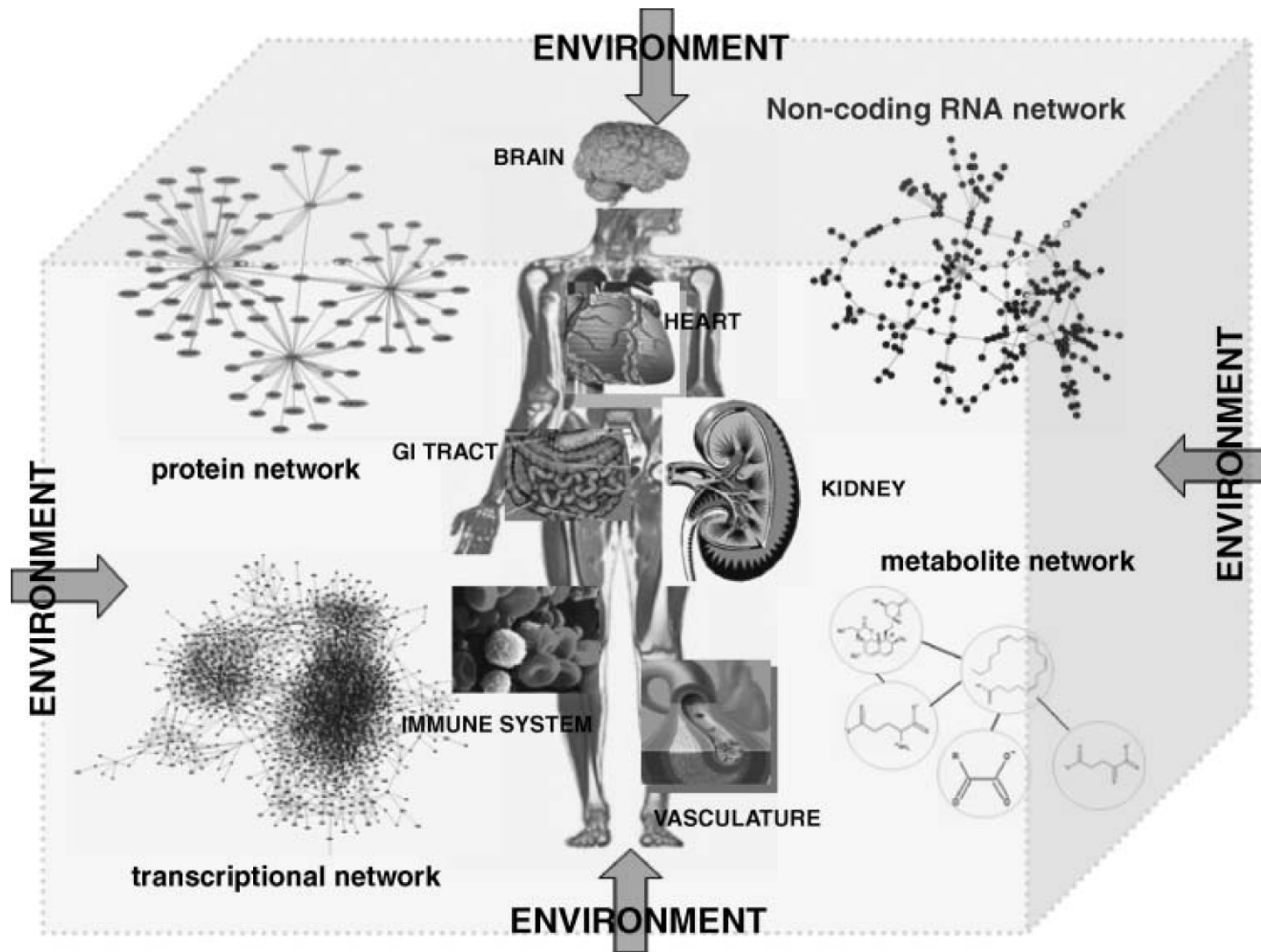
Examples of
4 functional
networks
driving the
development of
different
anatomical
structures in
the human
heart of a
37-day old
human embryo



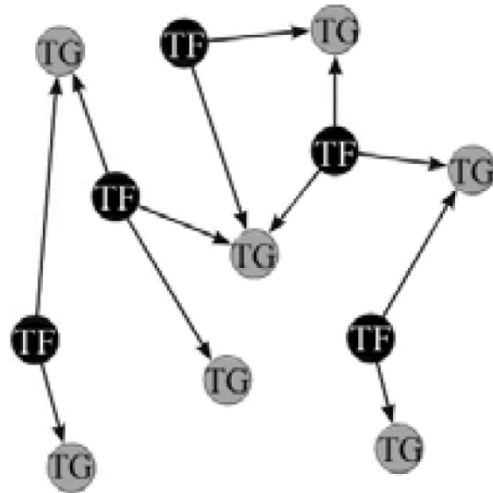
Lage, K. et. al (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular systems biology*, 6, 1, 1-9.



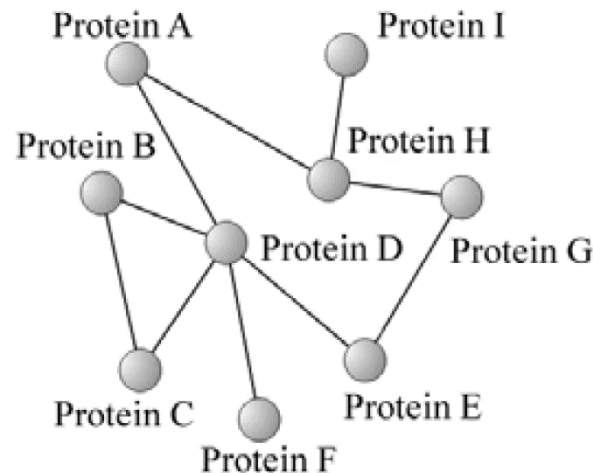
Lage et. al (2010)



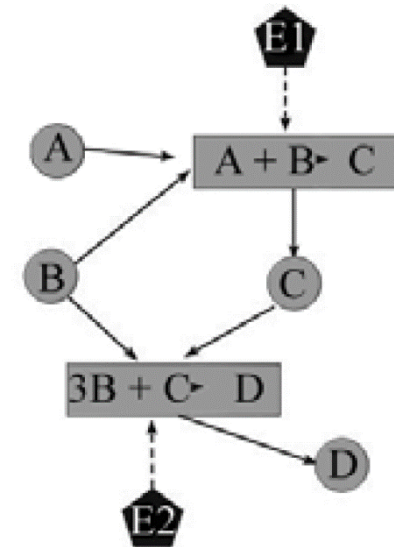
Schadt, E. E. & Lum, P. Y. (2006) Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *Journal of lipid research*, 47, 12, 2601-2613.



Transcriptional regulatory network with two components:
TF = transcription factor
TG = target genes
(TF regulates the transcription of TG)



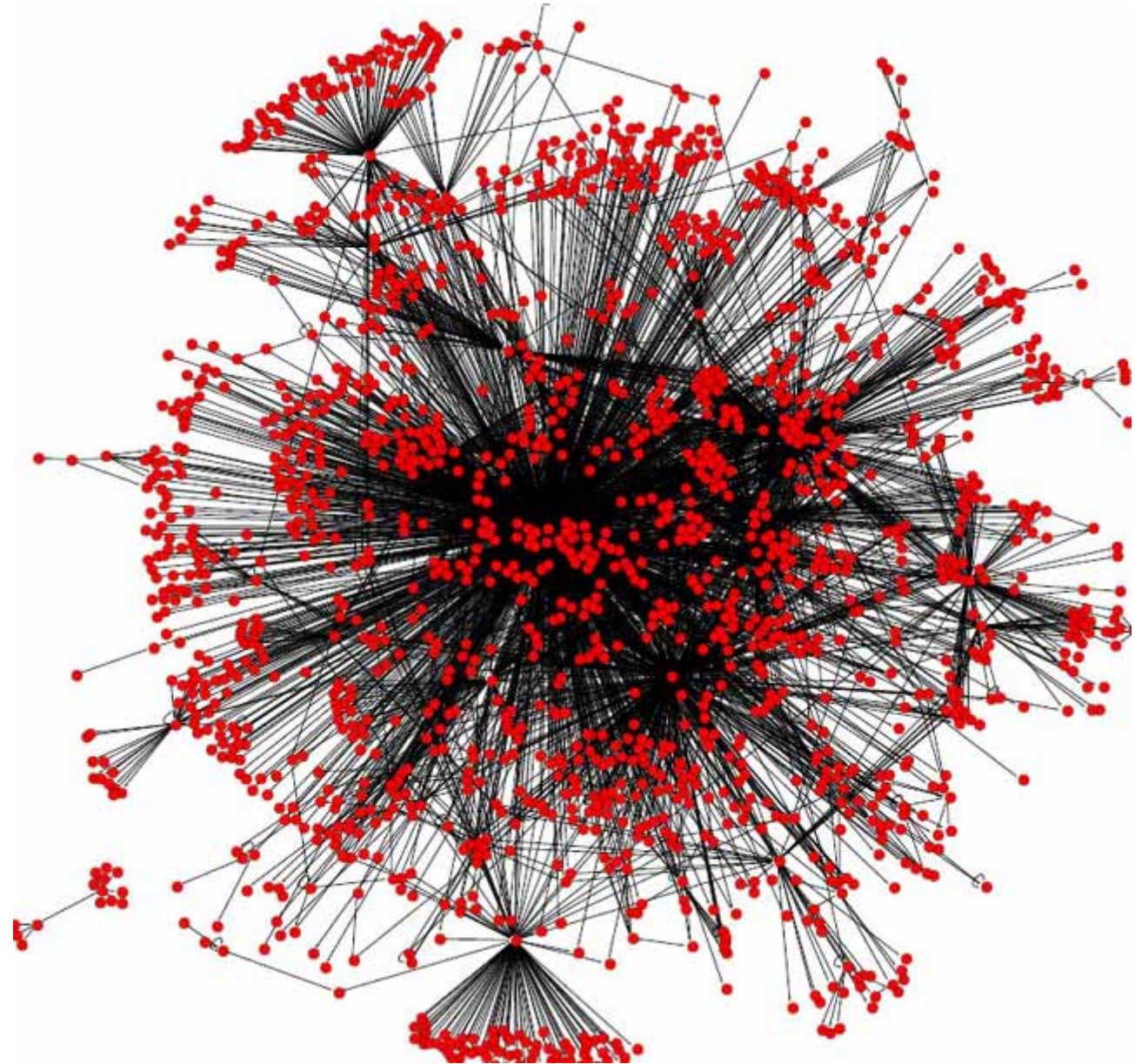
Protein-Protein interaction network

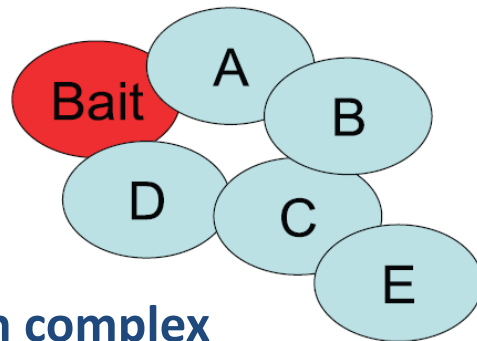
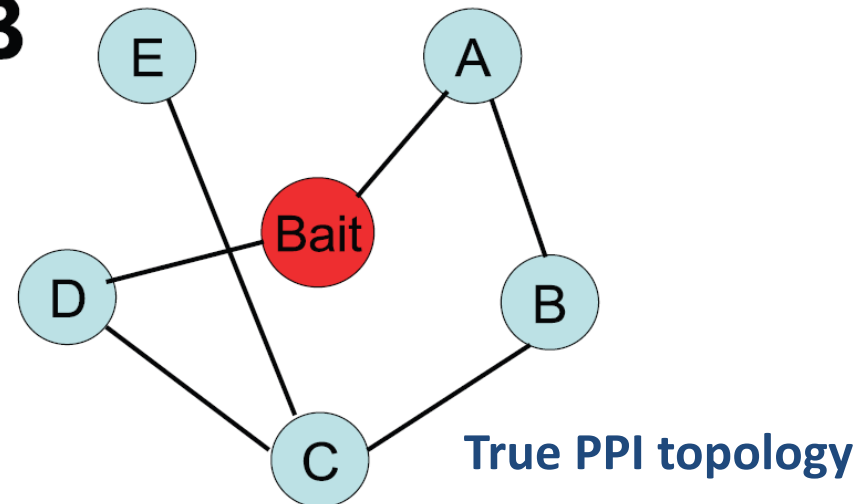
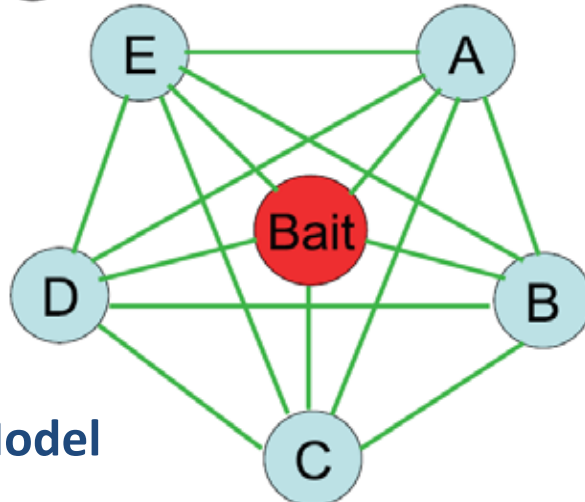
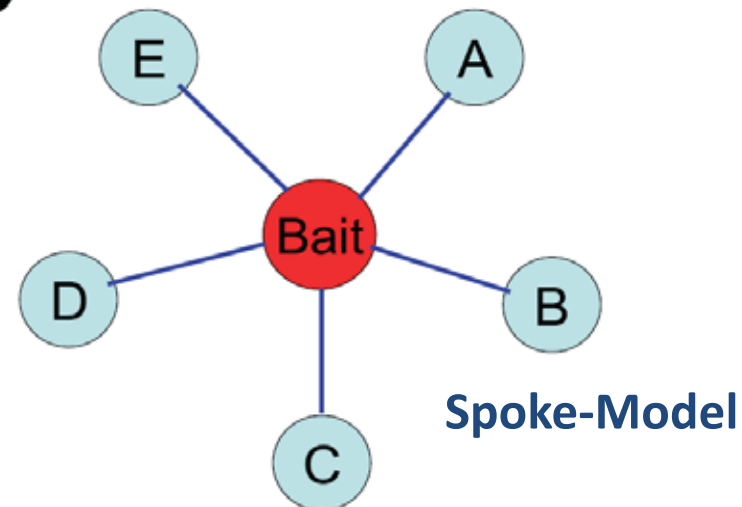


Metabolic network
(constructed considering the reactants, chemical reactions and enzymes)

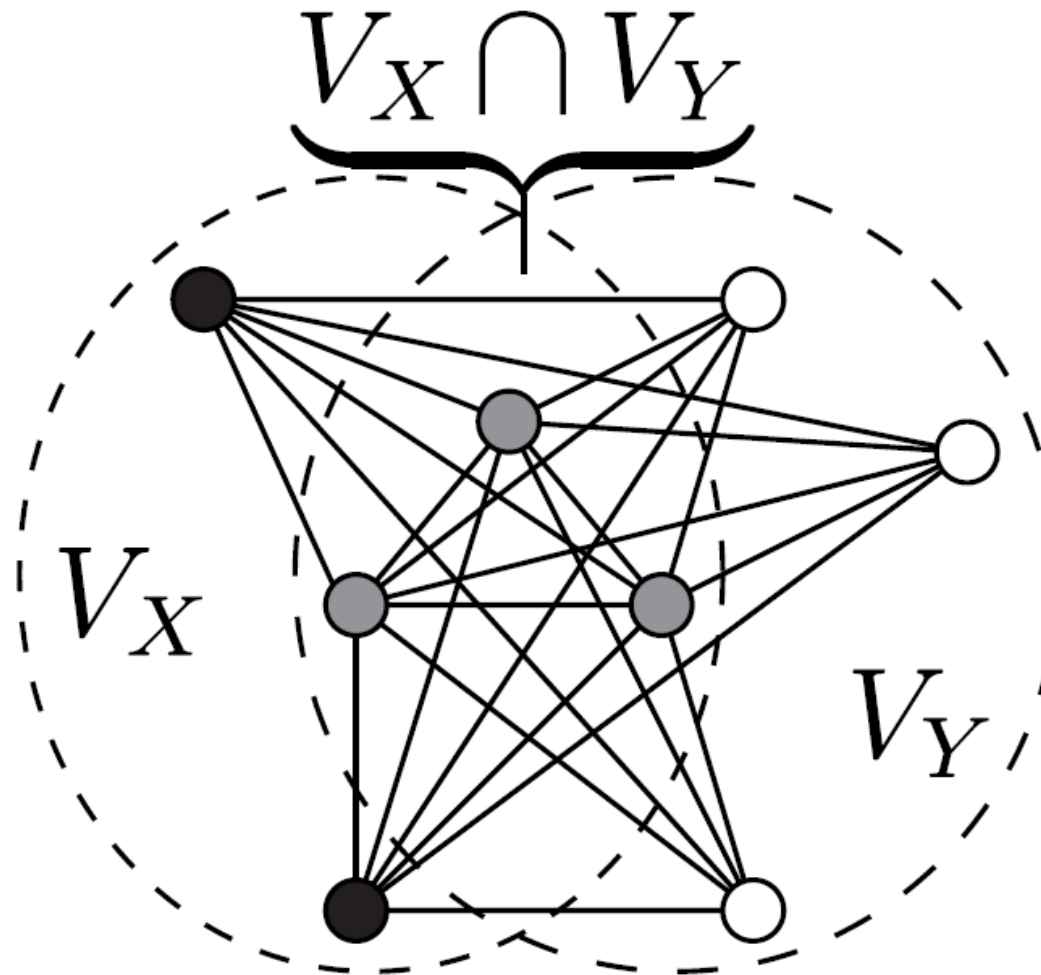
Costa, L. F., Rodrigues, F. A. & Cristino, A. S. (2008)
 Complex networks: the key to systems biology.
Genetics and Molecular Biology, 31, 3, 591–601.

Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Peralta-Gil, M., Peñaloza-Spínola, M. I., Martínez-Antonio, A., Karp, P. D. & Collado-Vides, J. 2006. The comprehensive updated regulatory network of *Escherichia coli* K-12. *BMC bioinformatics*, 7, (1), 5.



A**B****C****D**

Wang, Z. & Zhang, J. Z. (2007) In search of the biological significance of modular structures in protein networks. *PLoS Computational Biology*, 3, 6, 1011-1021.

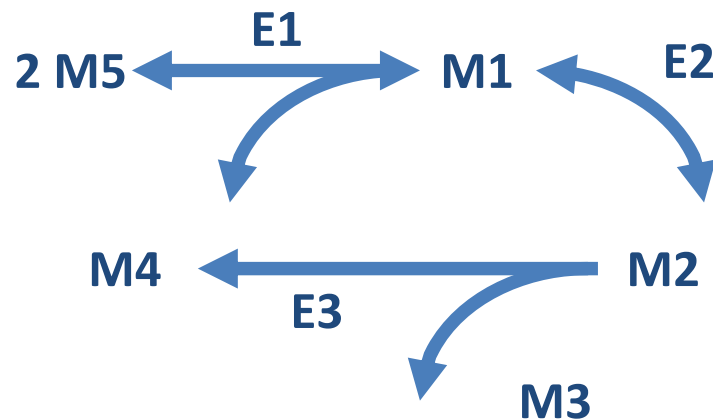


Boyen, P., Van Dyck, D., Neven, F., van Ham, R. C. H. J. & van Dijk, A. (2011) SLIDER: A Generic Metaheuristic for the Discovery of Correlated Motifs in Protein-Protein Interaction Networks. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 8, 5, 1344-1357.

Input: PPI-network $G = (V, E, \lambda)$, $\ell, d \in \mathbb{N}$, $d < \ell$
Output: $\{X^*, Y^*\}$ best correlated motif pair found in G

- 1: $\{X^*, Y^*\} \leftarrow \text{randomMotifPair}()$
- 2: $maxsup \leftarrow f(\{X^*, Y^*\}, G)$
- 3: $sup \leftarrow -\infty$
- 4: **while** $maxsup > sup$ **do**
- 5: $\{X, Y\} \leftarrow \{X^*, Y^*\}$
- 6: $sup \leftarrow maxsup$
- 7: **for all** $\{X', Y'\} \in N(\{X, Y\})$ **do**
- 8: **if** $f(\{X', Y'\}, G) > maxsup$ **then**
- 9: $\{X^*, Y^*\} \leftarrow \{X', Y'\}$
- 10: $maxsup \leftarrow f(\{X', Y'\}, G)$

Boyen et al. (2011)



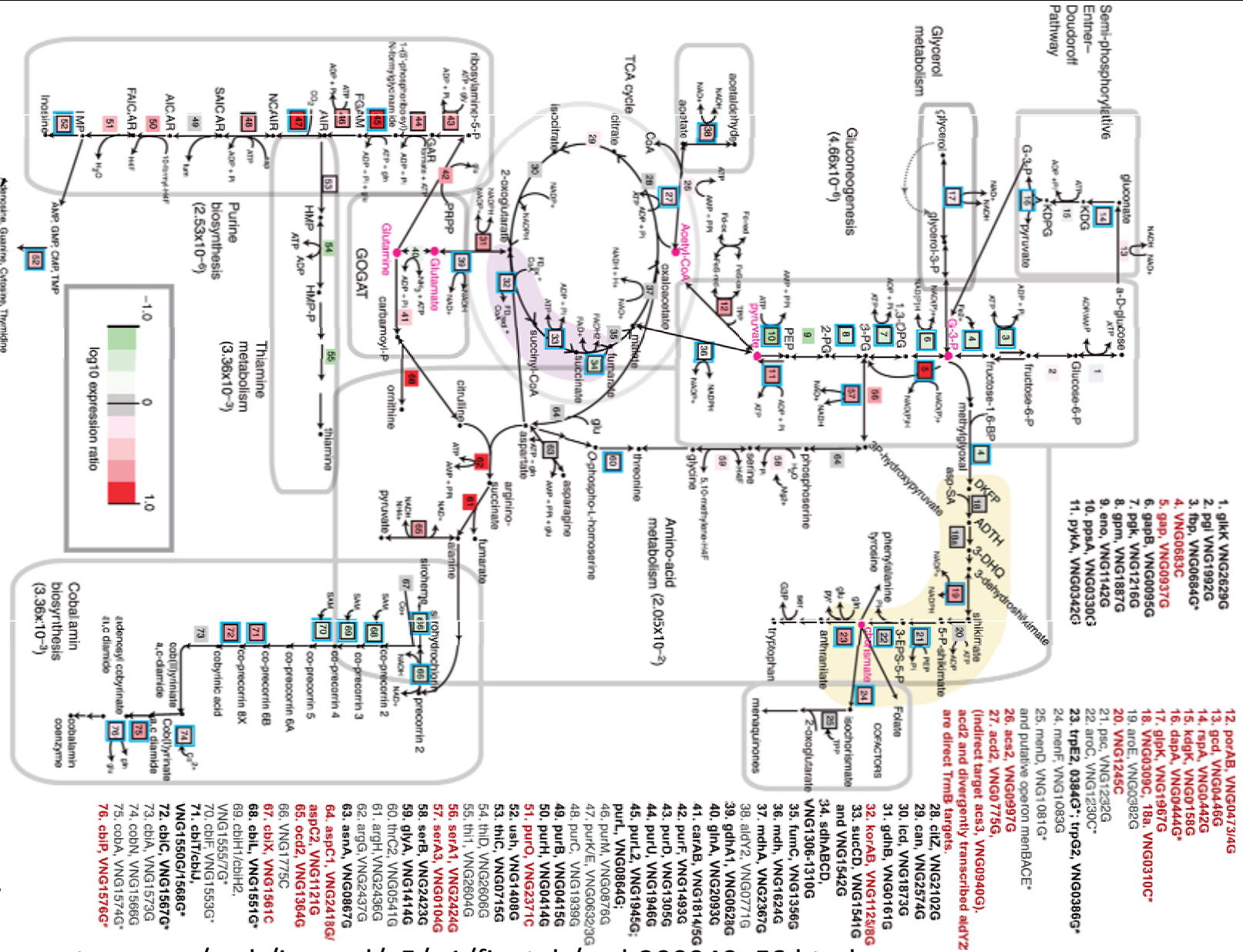
	M1	M2	M3	M4	M5
M1	0	1	0	1	1
M2	1	0	1	1	0
M3	0	0	0	0	0
M4	1	0	0	0	0
M5	1	0	0	0	0

Matrix contains many sparse elements - In this case it is computationally more efficient to represent the graph as an adjacency list

M1	M2
M1	M4
M1	M5
M2	M1
M2	M3
M2	M4
M4	M1
M5	M1

Hodgman, C. T., French, A. & Westhead, D. R. (2010) *Bioinformatics. Second Edition*. New York, Taylor & Francis.

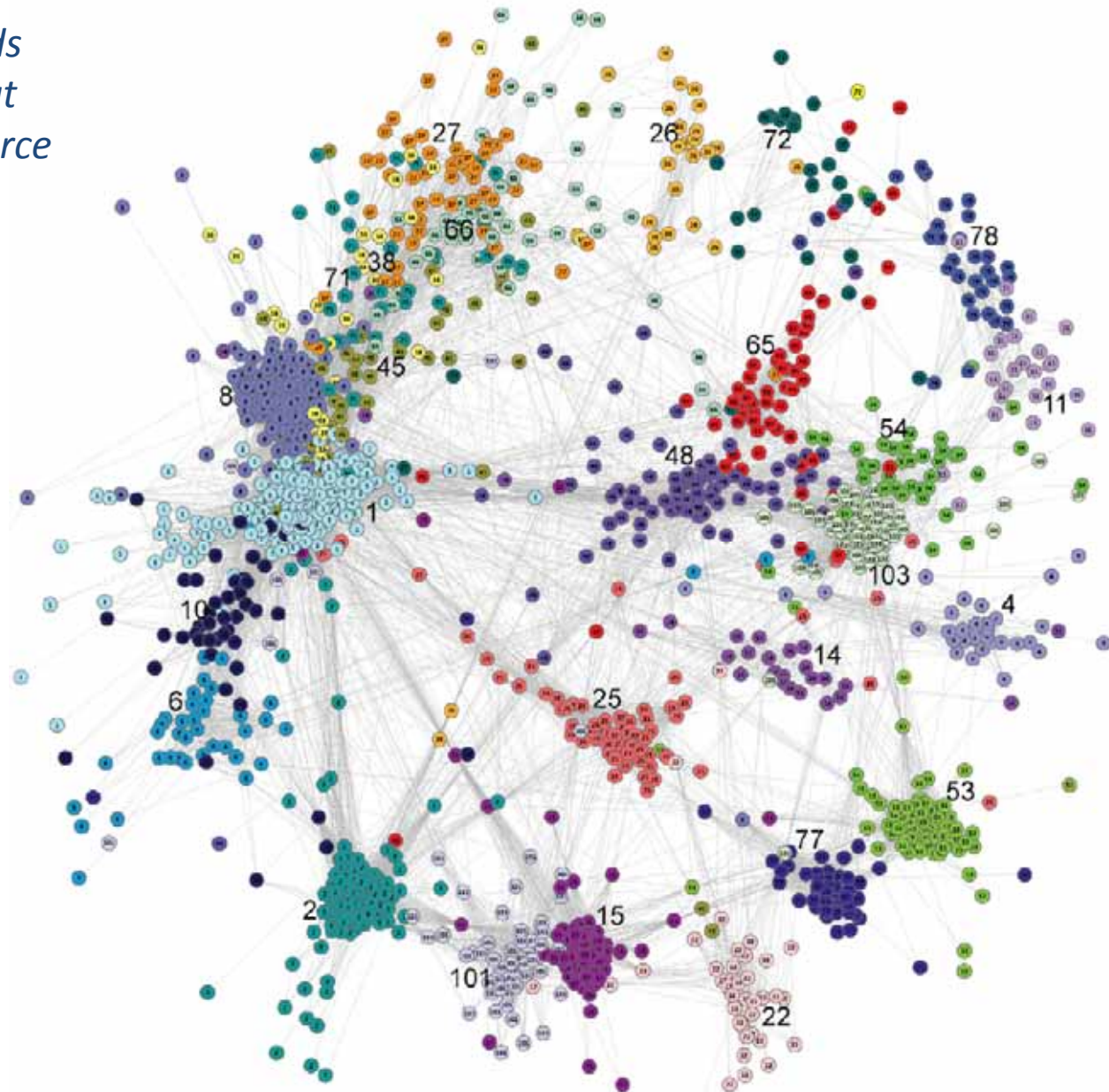
Schmid, A. K.,
Reiss, D. J.,
Pan, M., Koide,
T. & Baliga, N.
S. (2009) A
single
transcription
factor
regulates
evolutionarily
diverse but
functionally
linked
metabolic
pathways in
response to
nutrient
availability.
*Molecular
Systems
Biology*, 5, 1-9.

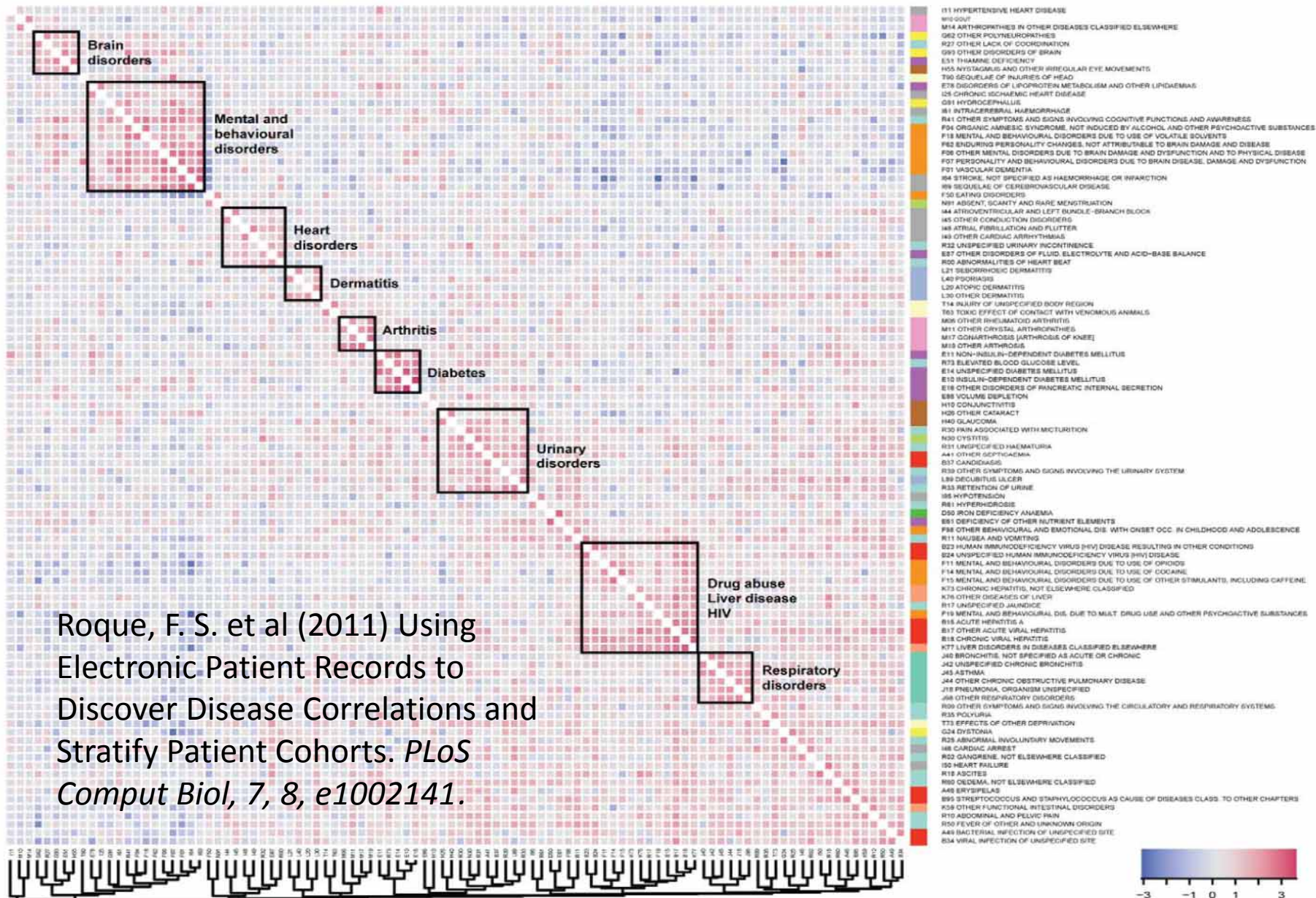


http://www.nature.com/msb/journal/v5/n1/fig_tab/msb200940_F6.html

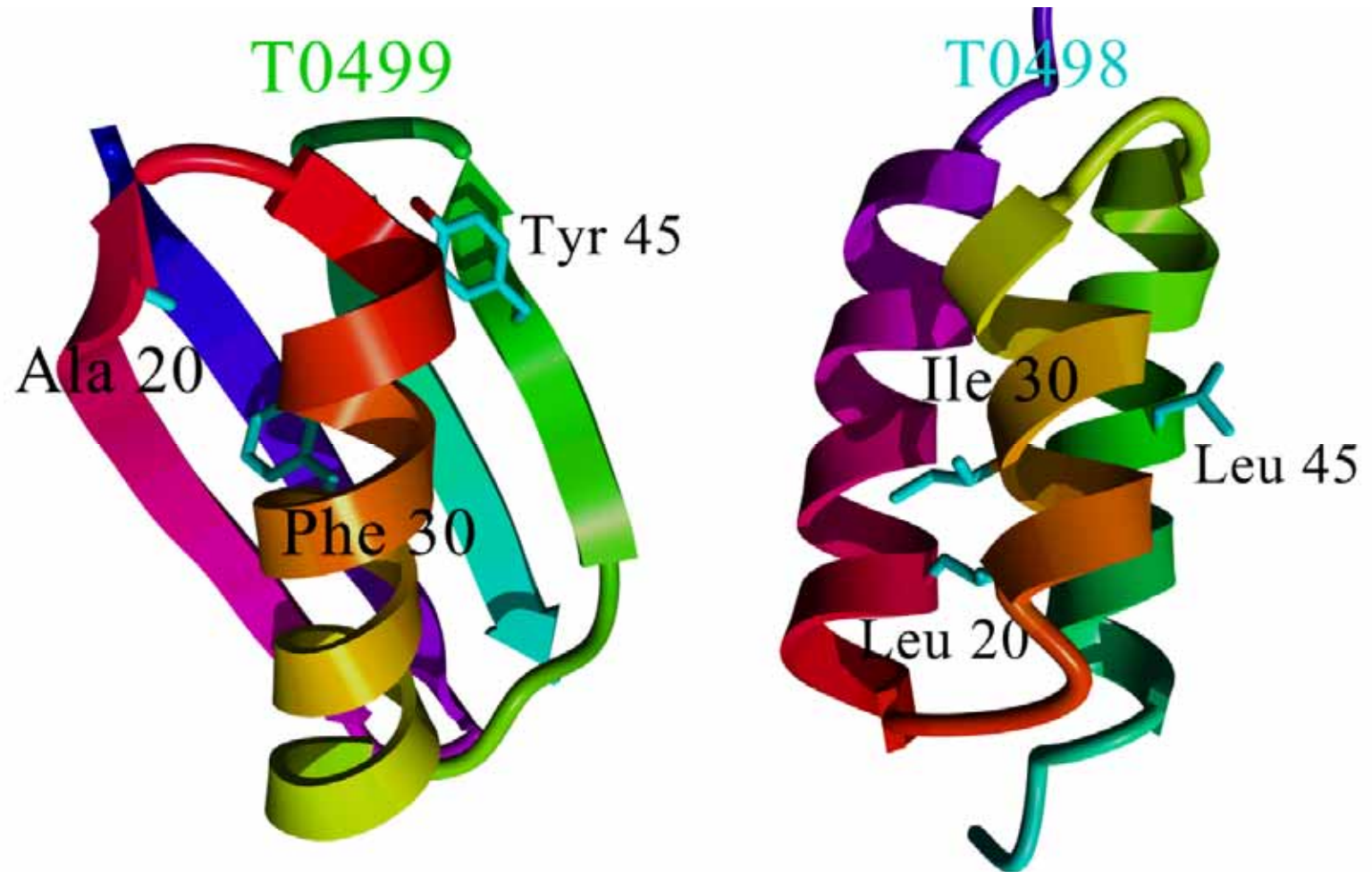
Electronic patient records remain a unexplored, but potentially rich data source for example to discover correlations between diseases.

Roque, F. S., Jensen, P. B., Schmock, H., Dalgaard, M., Andreatta, M., Hansen, T., Søbey, K., Bredkjær, S., Juul, A., Werge, T., Jensen, L. J. & Brunak, S. (2011) Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. *PLoS Computational Biology*, 7, 8, e1002141.



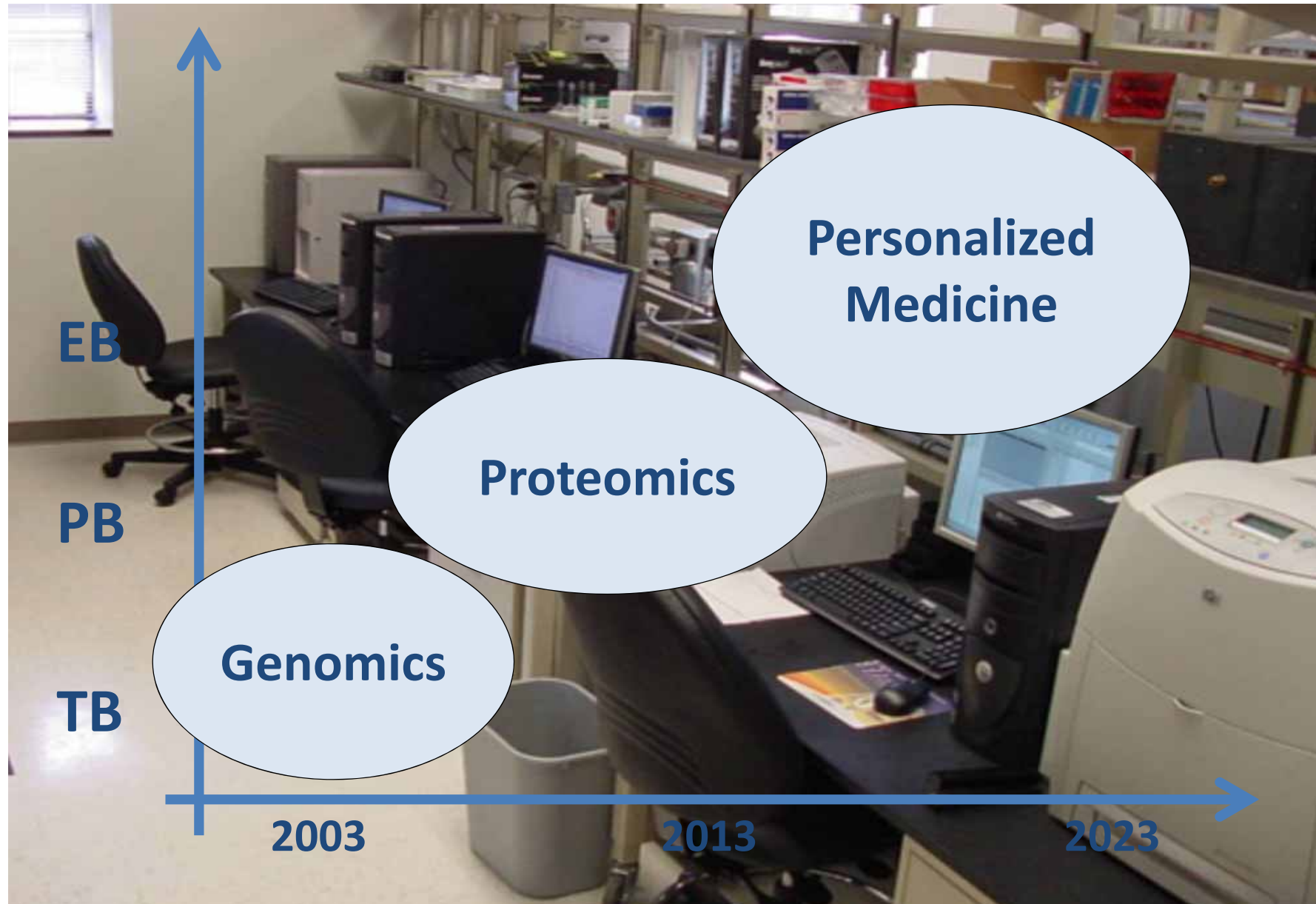


He, Y., Chen, Y.,
Alexander, P.,
Bryan, P. N. &
Orban, J. (2008)
NMR structures of
two designed
proteins with high
sequence identity
but different fold
and function.
Proceedings of the
National Academy
of Sciences, 105,
38, 14412.



T0499	TTYKL I LNLKQAKEEA I KEAVDAGTAEKY FKL I ANAKTVEGWWTYKDE I KTFTVTE
	X X X
T0498	TTYKL I LNLKQAKEEA I KELVDAGTAEKY I KL I ANAKTVEGWTLKDE I KTFTVTE

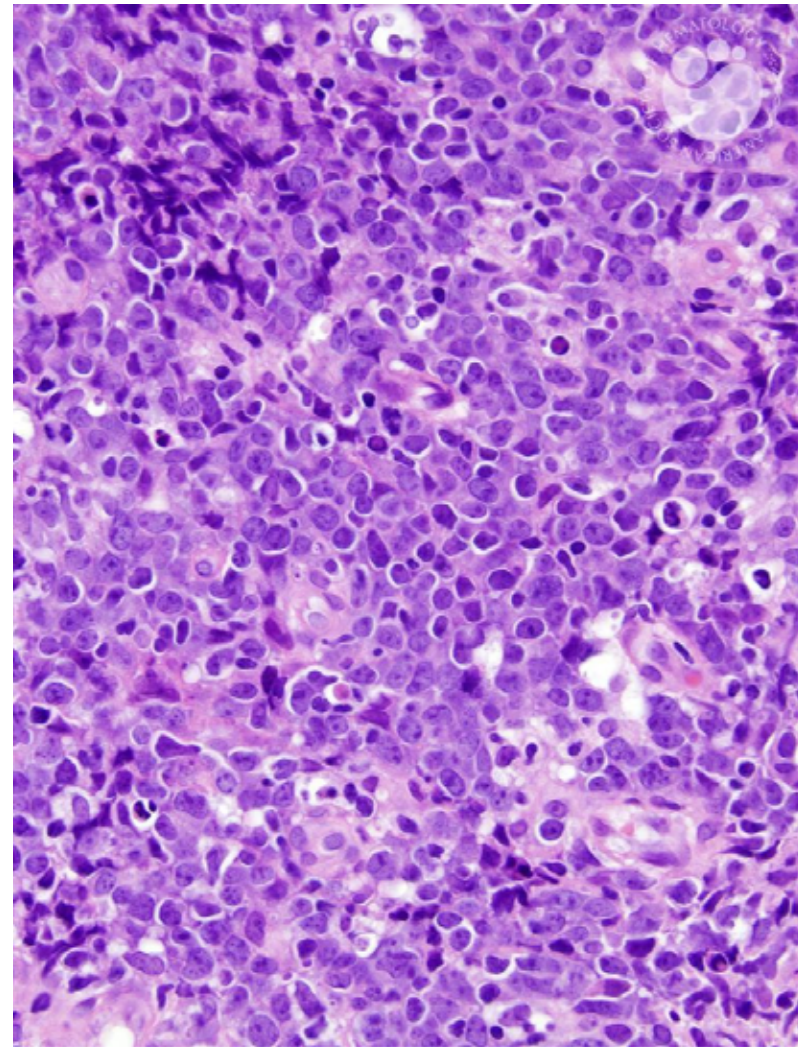
- Homology modeling is a knowledge-based prediction of protein structures.
- In homology modeling a protein sequence with an unknown structure (the target) is aligned with one or more protein sequences with known structures (the templates).
- The method is based on the principle that homologue proteins have similar structures.
- **Homology modeling will be extremely important to personalized and molecular medicine in the future.**




04 Machine Learning on Graphs Examples

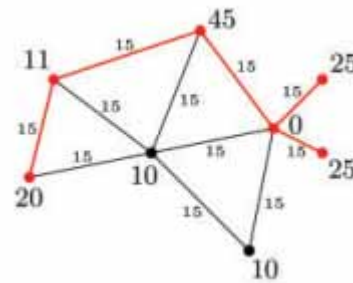
The two main forms of lymphoma are Hodgkin lymphoma and non-Hodgkin lymphoma (NHL). Lymphoma occurs when cells of the immune system called lymphocytes, a type of white blood cell, grow and multiply uncontrollably. Cancerous lymphocytes can travel to many parts of the body, including the lymph nodes, 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



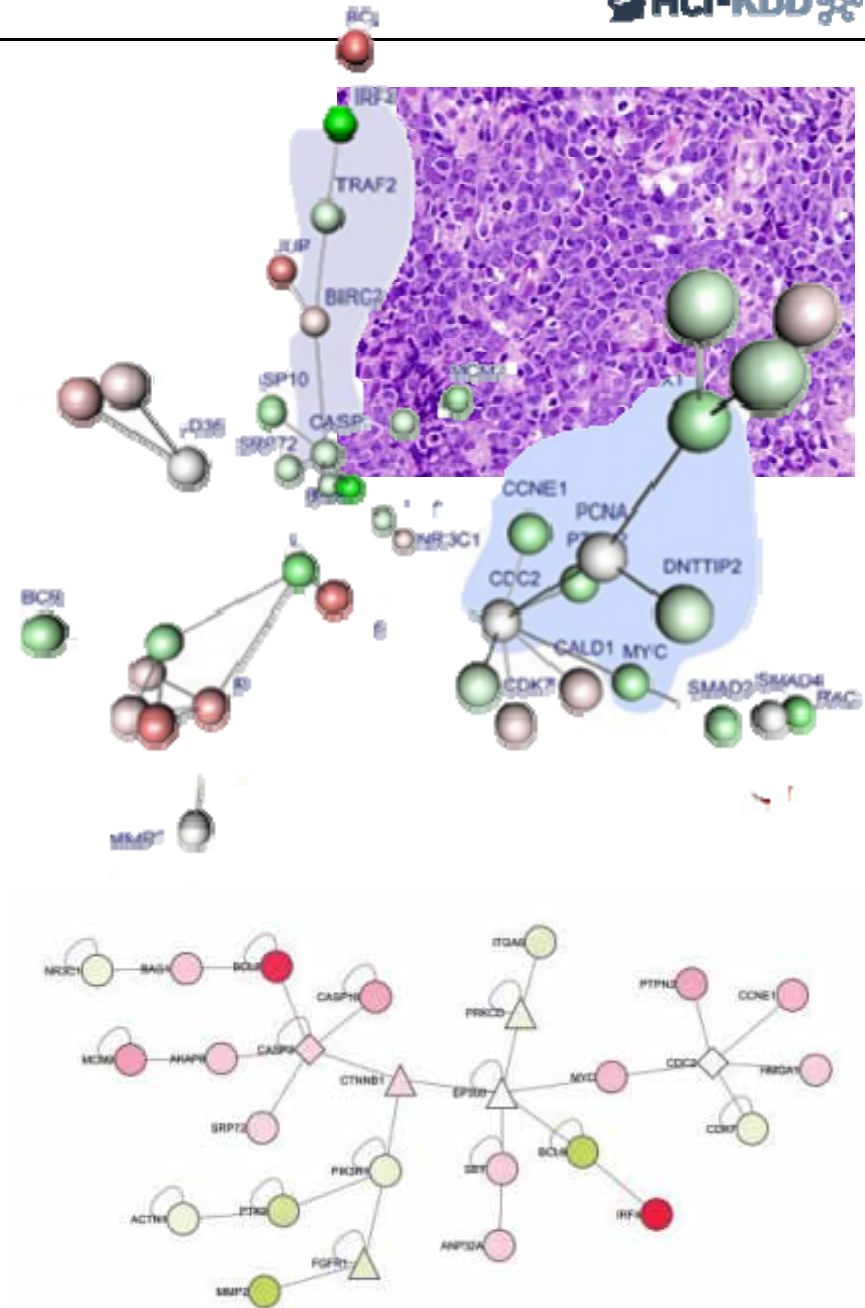
<http://imagebank.hematology.org/>

- 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.

Holzinger Group, hci-kdd.org



Subscribe Recommend Publish Editorial Boards Testimonials About Welcome, [Learn more about access or create an account.](#)

jove Search by keywords, for example: "stem cells" Advanced Search Sign In

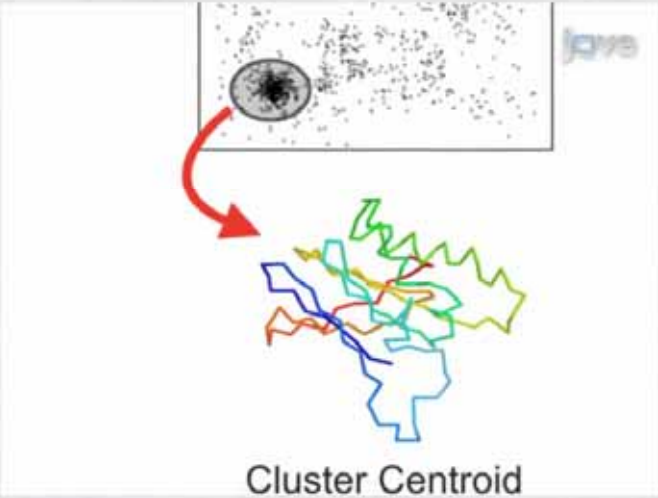
8 A Protocol for Computer-Based Protein Structure and Function Prediction

Ambresh Roy^{1,2}, Dong Xu¹, Jonathan Poisson¹, Yang Zhang^{1,2}

¹Center for Computational Medicine and Bioinformatics, **University of Michigan**, ²Center for Bioinformatics and Department of Molecular Bioscience, **University of Kansas**

Article Downloads Comments Metrics

27



Cluster Centroid

0:05 Title
2:21 Running the I-TASSER Server
3:37 Structure Analysis
5:58 LOMETS Target Template Alignment
7:30 Structural Analogs in PDB and Enzyme Commission Number Prediction
9:20 Gene Ontology (GO) Term and Protein-ligand Bind site Predictions
12:05 Representative I-TASSER Results
15:43 Conclusion

This article is Open Access.

Recommend JoVE to Your Librarian

Related Videos

- The ITS2 Database Published 3/12/2012
- Analyzing and Building Nucleic Acid Structures... Published 4/26/2013
- Protein WISDOM: A Workbench for In... Published 7/25/2013
- Optimization of Synthetic Proteins... Published 7/14/2015

Translate text to:
Choose Language...

Summary

Guidelines for computer based structural and functional characterization of protein using the I-TASSER pipeline is described. Starting from query protein sequence, 3D models are generated using multiple threading alignments and iterative structural assembly simulations. Functional inferences are thereafter drawn

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

Nodes: proteins

Links: physical interactions (binding)

Puzzling pattern:

Hubs tend to link to small degree nodes.

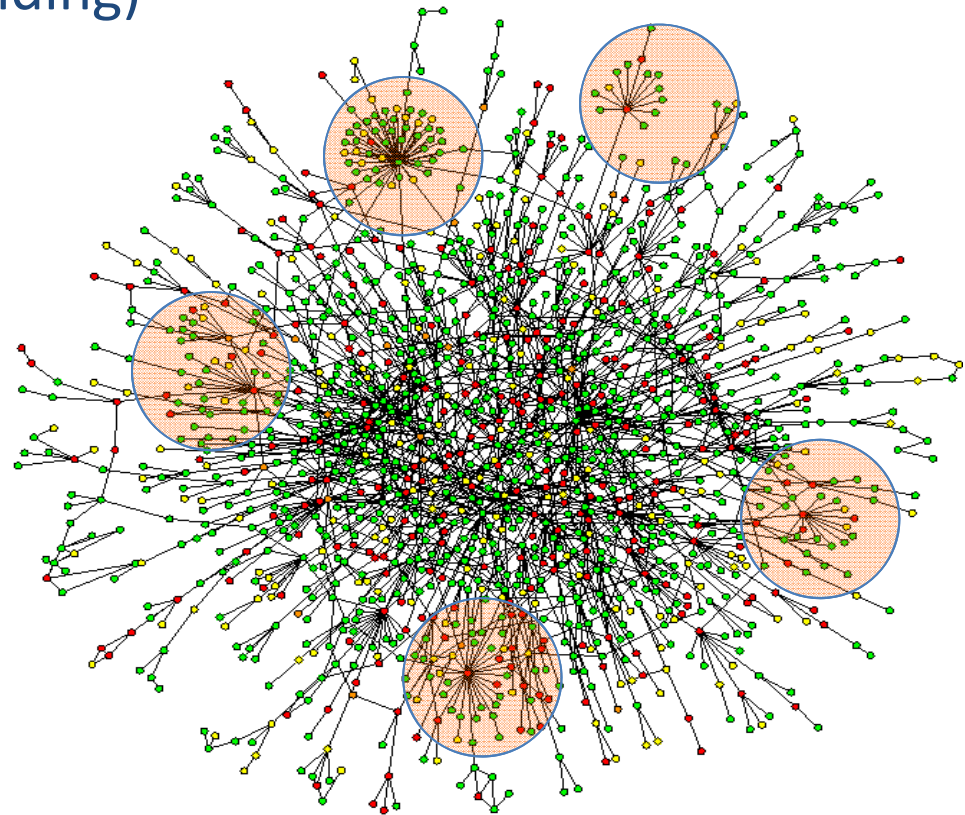
Why is this puzzling?

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

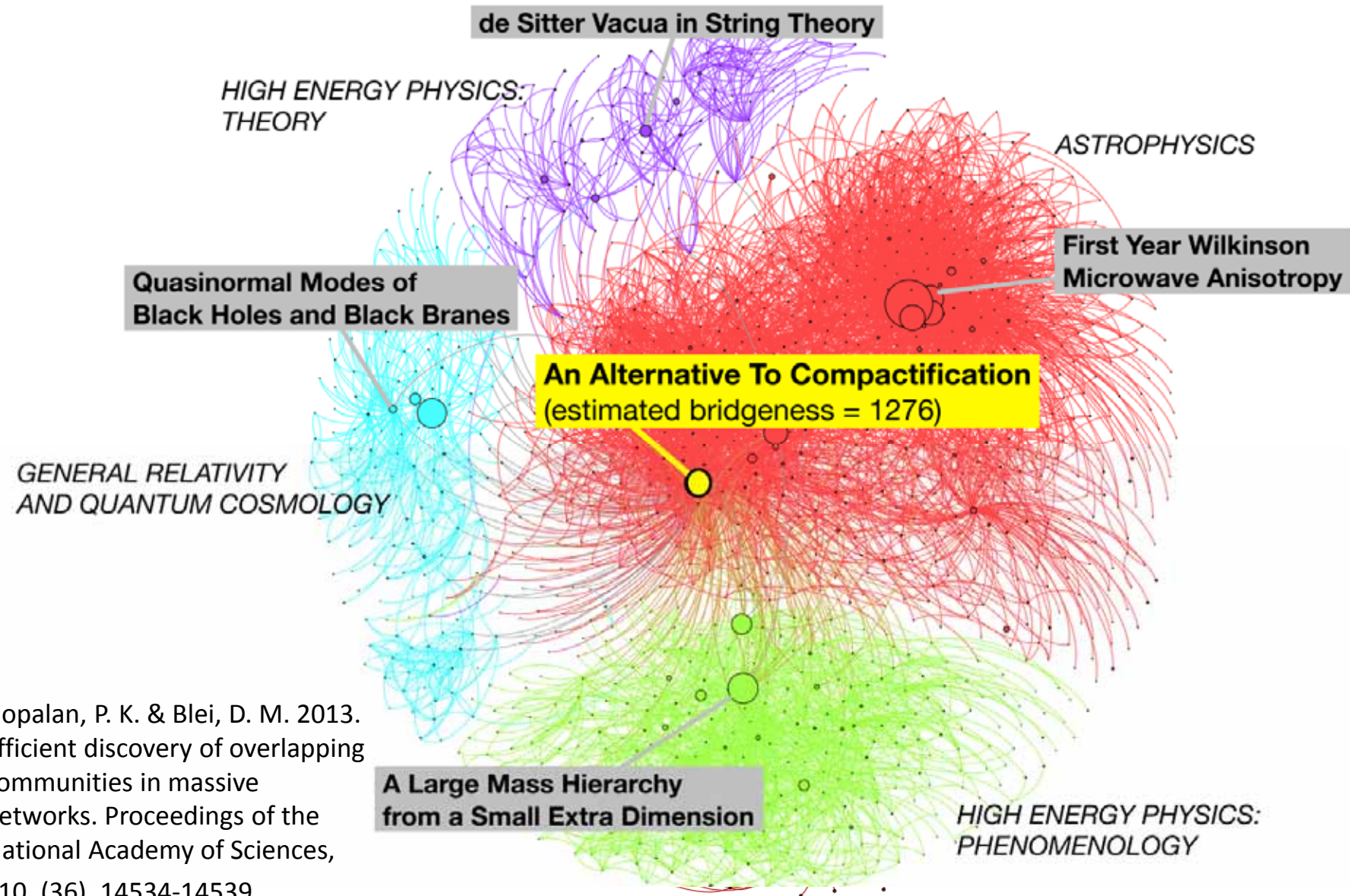
$$p_{kk'} = \frac{kk'}{2L}$$

$$k \approx 50, k' = 13, N = 1,458, L = 1746$$

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

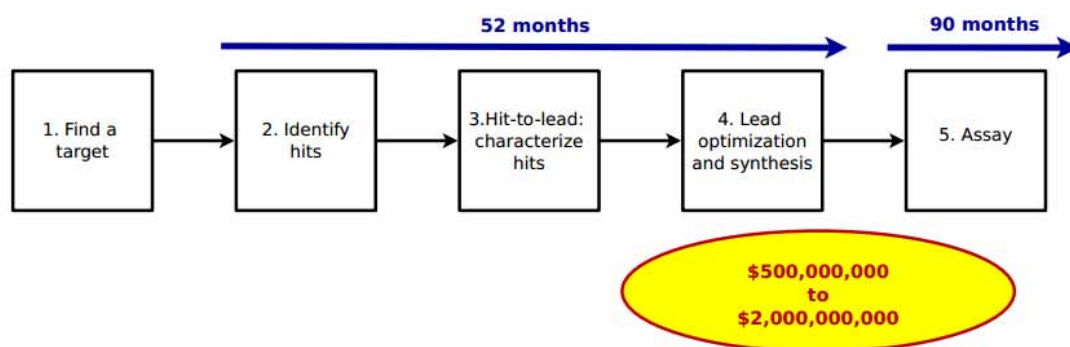


Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. 2001. Lethality and centrality in protein networks. Nature, 411, (6833), 41-42.



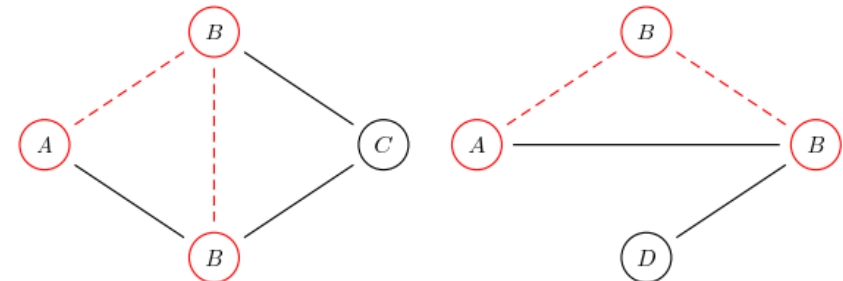
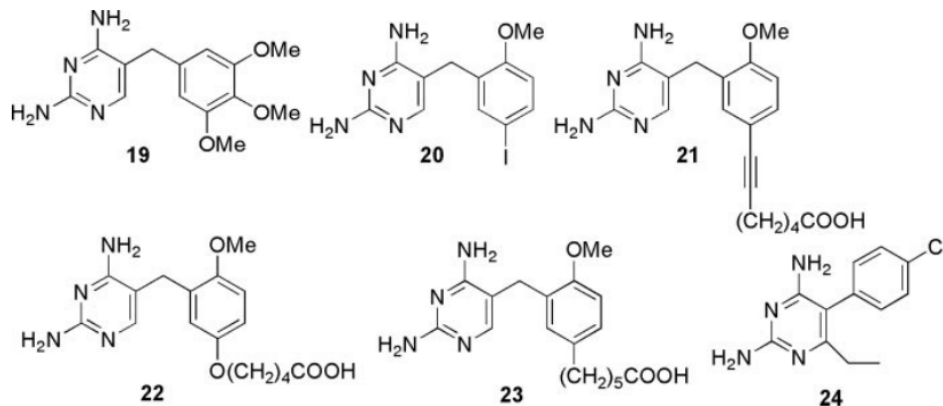
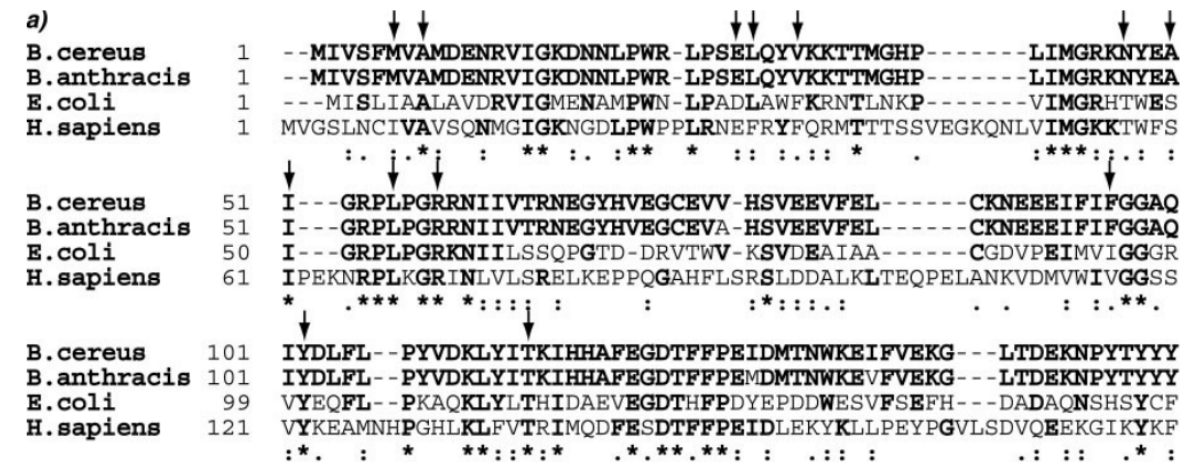
Gopalan, P. K. & Blei, D. M. 2013. Efficient discovery of overlapping communities in massive networks. Proceedings of the National Academy of Sciences, 110, (36), 14534-14539.

- 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



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

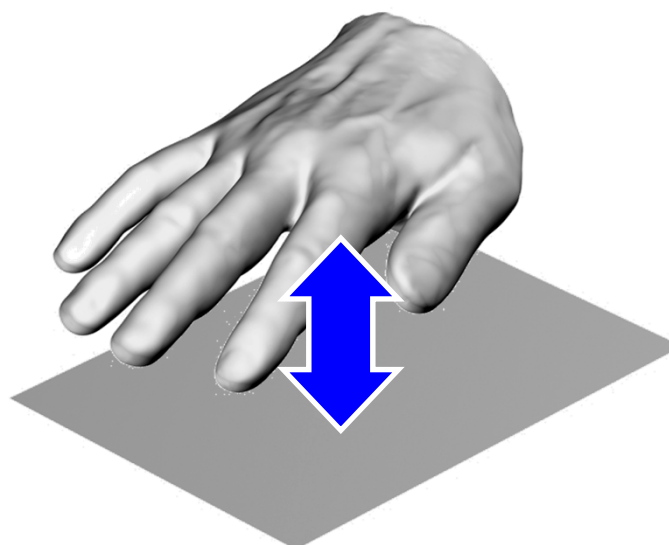
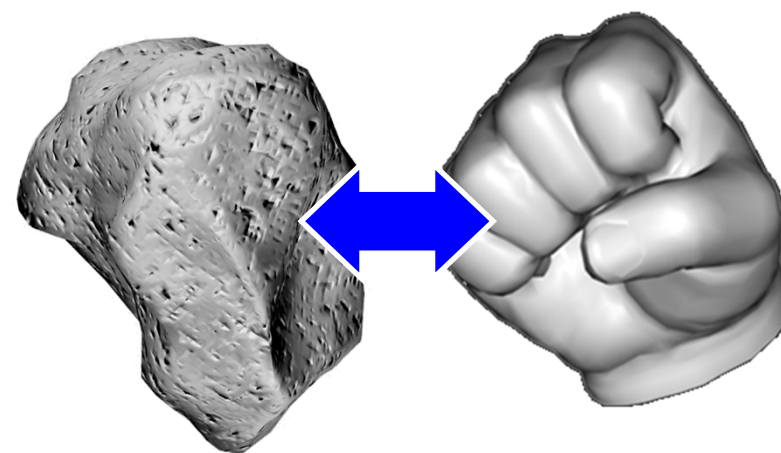
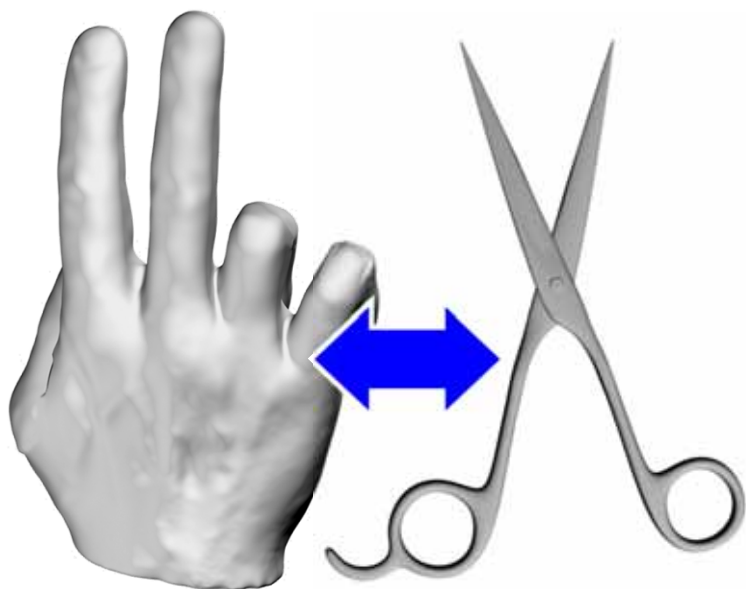
Joska, T. M. & Anderson, A. C. 2006. Structure-activity relationships of *Bacillus cereus* and *Bacillus anthracis* dihydrofolate reductase: toward the identification of new potent drug leads. *Antimicrobial agents and chemotherapy*, 50, 3435-3443.

- 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

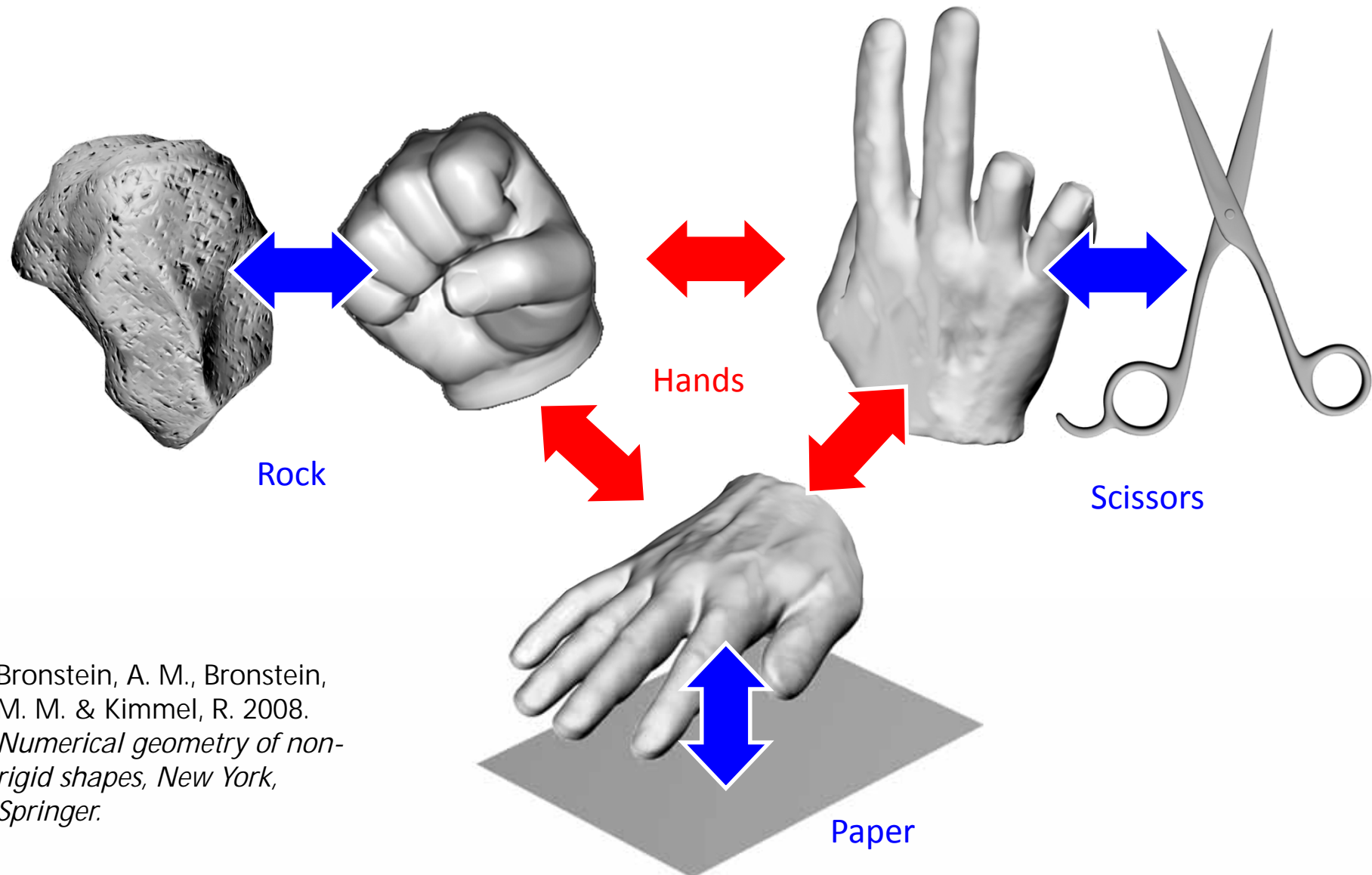
05 Digression: What is similarity?



Image credit to Eamonn Keogh (2008)



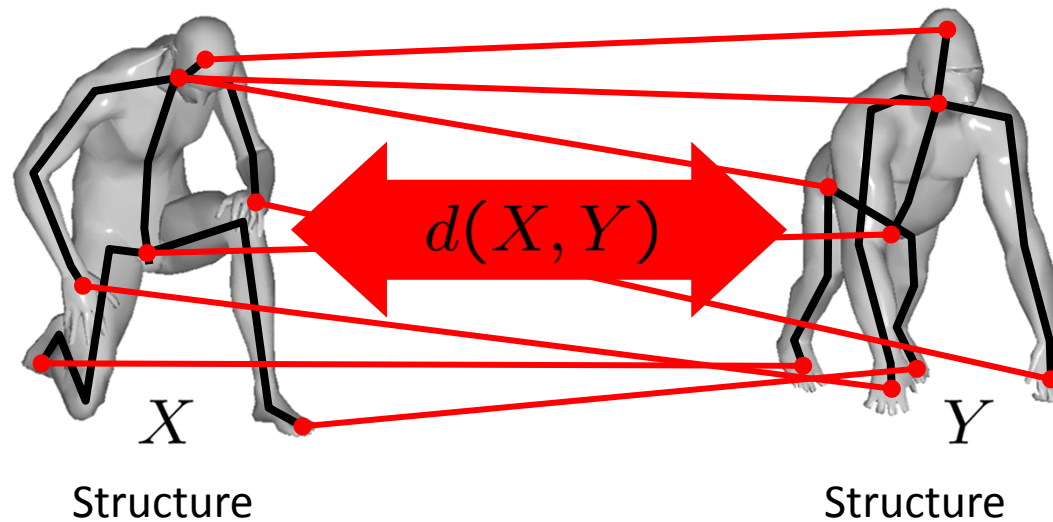
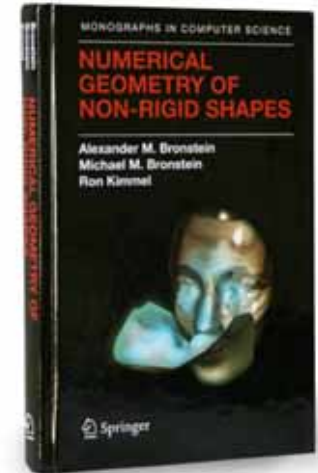
Bronstein, A. M., Bronstein,
M. M. & Kimmel, R. 2008.
*Numerical geometry of non-
rigid shapes*, New York,
Springer.



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

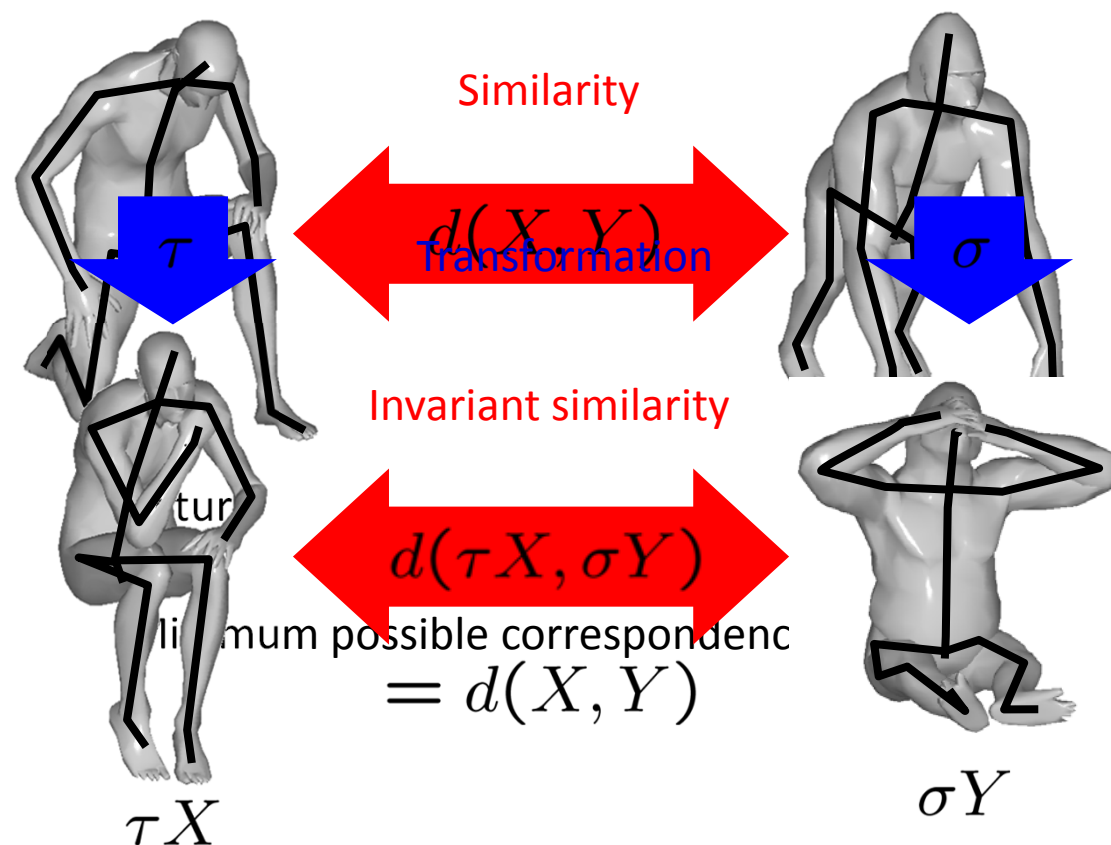
Bronstein, A. M., Bronstein, M. M. & Kimmel, R. 2008. *Numerical geometry of non-rigid shapes*, New York, Springer.

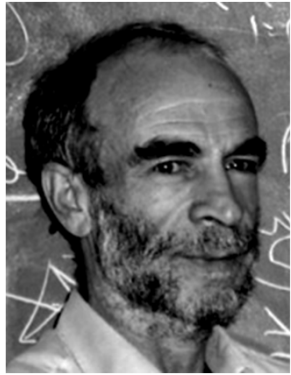
<http://www.inf.usi.ch/bronstein/>



Correspondence quality = structure similarity
(distortion)

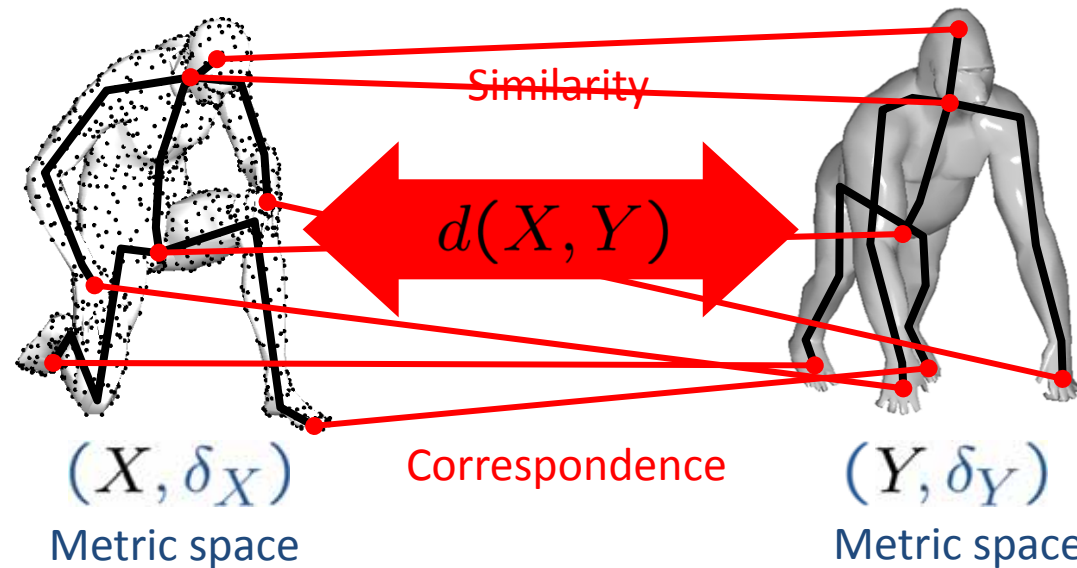
Minimum possible correspondence distortion





Michail Gromov
(1943-)

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



Felix Hausdorff
(1868-1942)

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

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

Discrete optimization over correspondences is NP hard !

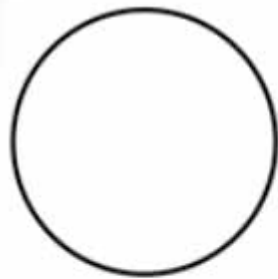


Enrico Betti
(1823-1892)

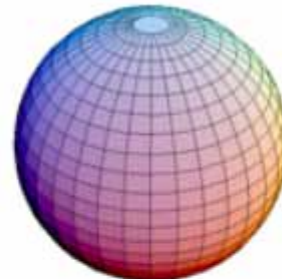
Counts the number of “i-dimensional holes”
 b_i is the “i-th Betti number”



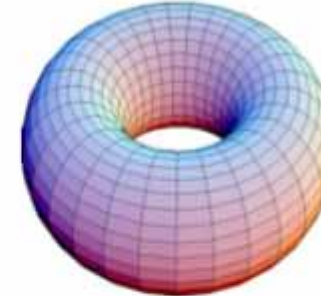
Emmy Noether
(1882-1935)



$$b_1=1$$
$$b_2=0$$



$$b_1=0$$
$$b_2=1$$

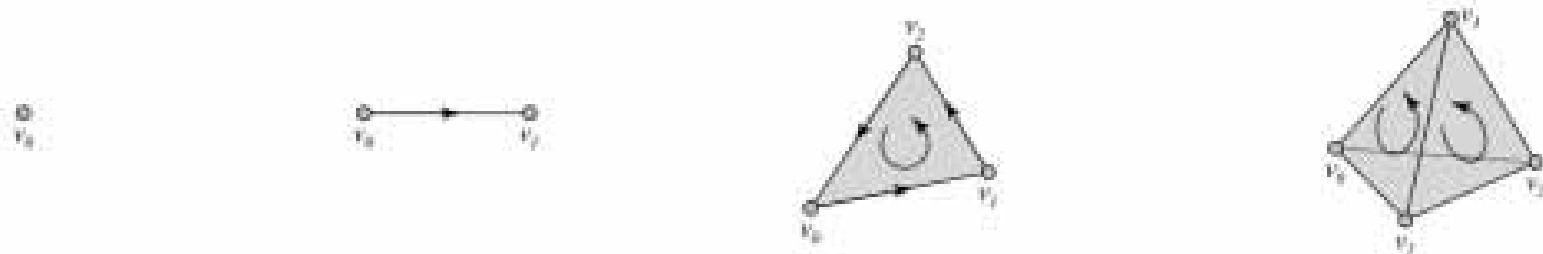


$$b_1=2$$
$$b_2=1$$

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.

- 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 $f: M \rightarrow \mathbb{R}^d$ 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 (Ω, \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 ?



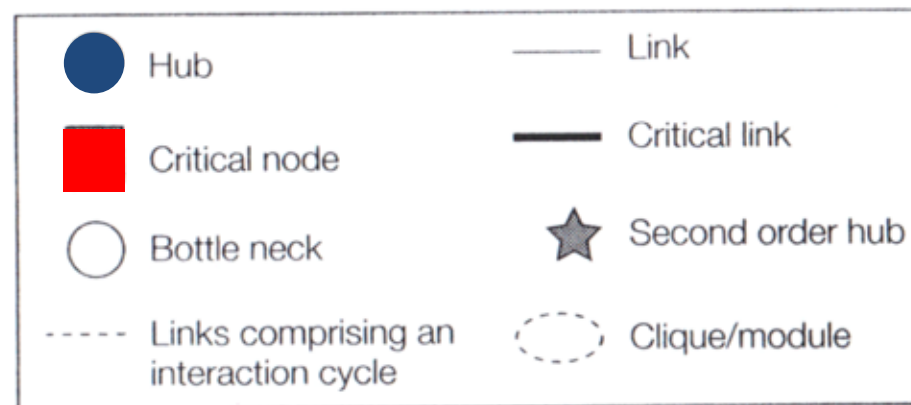
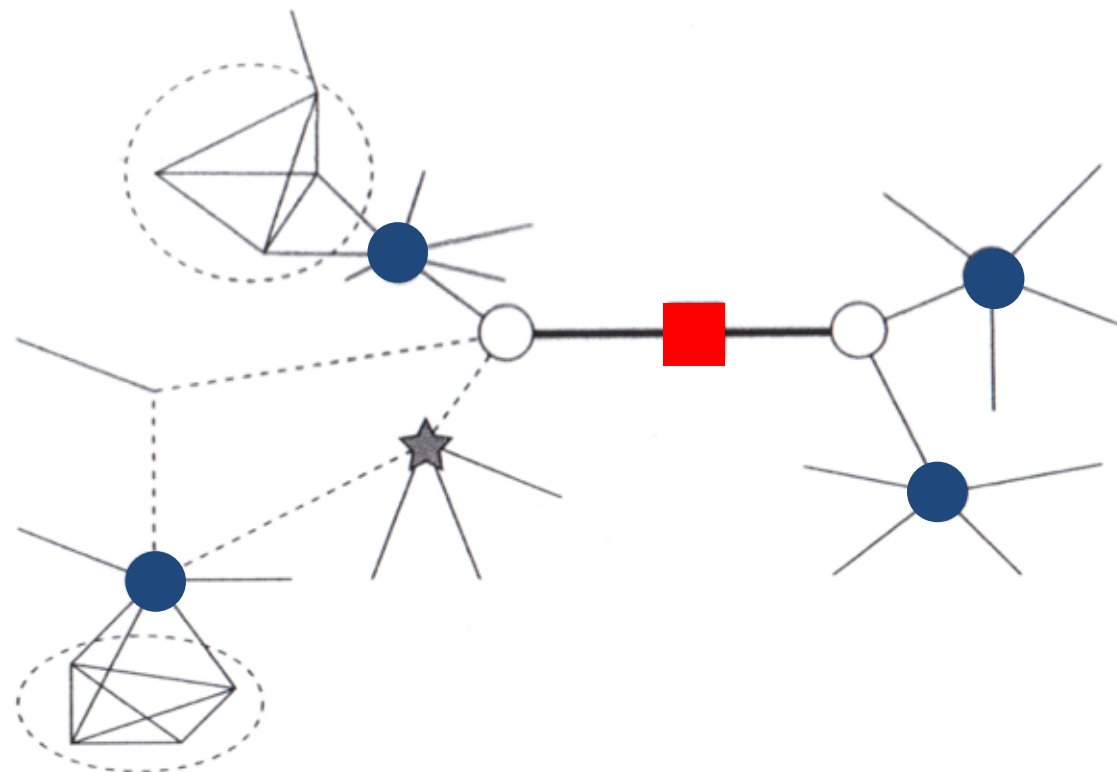
- Mega Problem: To date none of our known methods, algorithms and tools scale to the massive amount and dimensionalities of data we are confronted in practice;
- we need much more research efforts towards making computational topology successful as a general method for data mining and knowledge discovery

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.

06 Review of basic concepts, metrics and measures

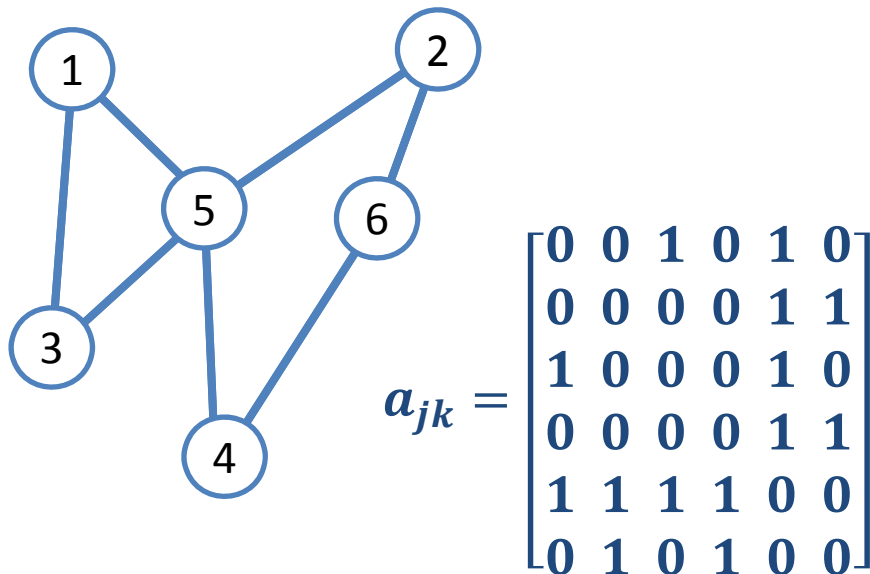
- In order to understand complex biological systems, the three following key concepts need to be considered:
- (i) **emergence**, the discovery of links between elements of a system because the study of individual elements such as genes, proteins and metabolites is insufficient to explain the behavior of whole systems;
- (ii) **robustness**, biological systems maintain their main functions even under perturbations imposed by the environment; and
- (iii) **modularity**, vertices sharing similar functions are highly connected.
- Network theory can largely be applied for biomedical informatics, because many tools are already available

$G(V, E)$ Graph
 V ... vertex
 E ... edge $\{a, b\}$
 $a, b \in V; a \neq b$

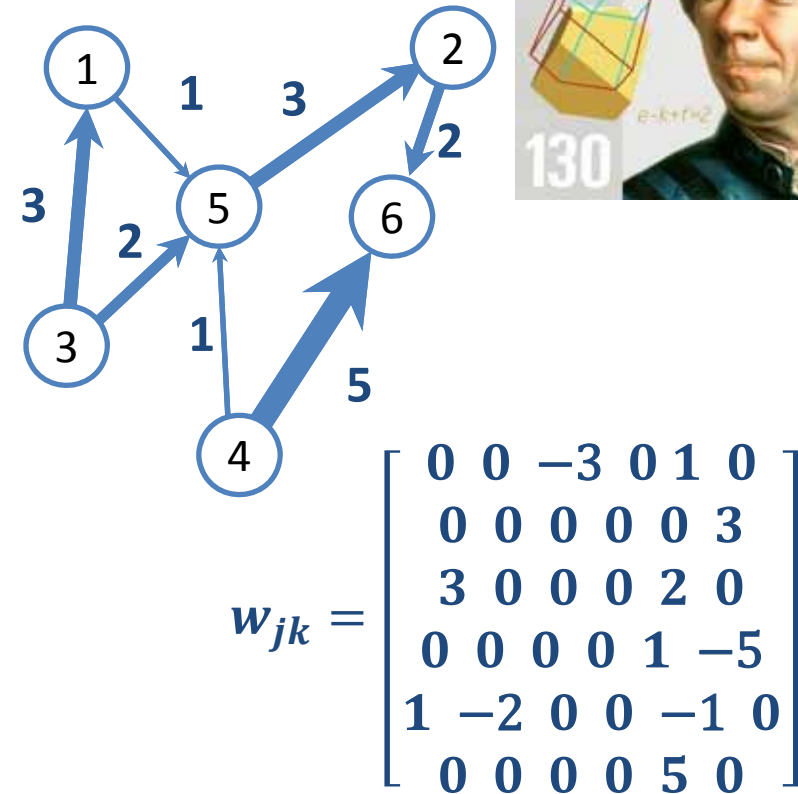


Hodgman, C. T.,
 French, A. &
 Westhead, D. R.
 (2010) *Bioinformatics*.
 Second Edition. New
 York, Taylor & Francis.

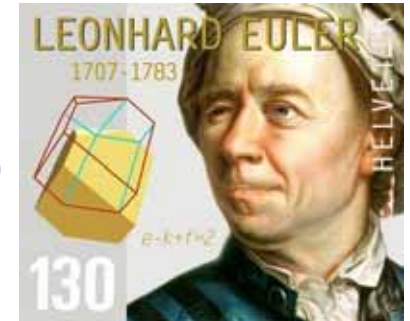
Adjacency (ə-ˈjā-sən(t)-sē) Matrix $A = (a_{jk})$
$$a_{jk} = \begin{cases} 1, & \text{if } \{j, k\} \in E \\ 0, & \text{otherwise} \end{cases}$$



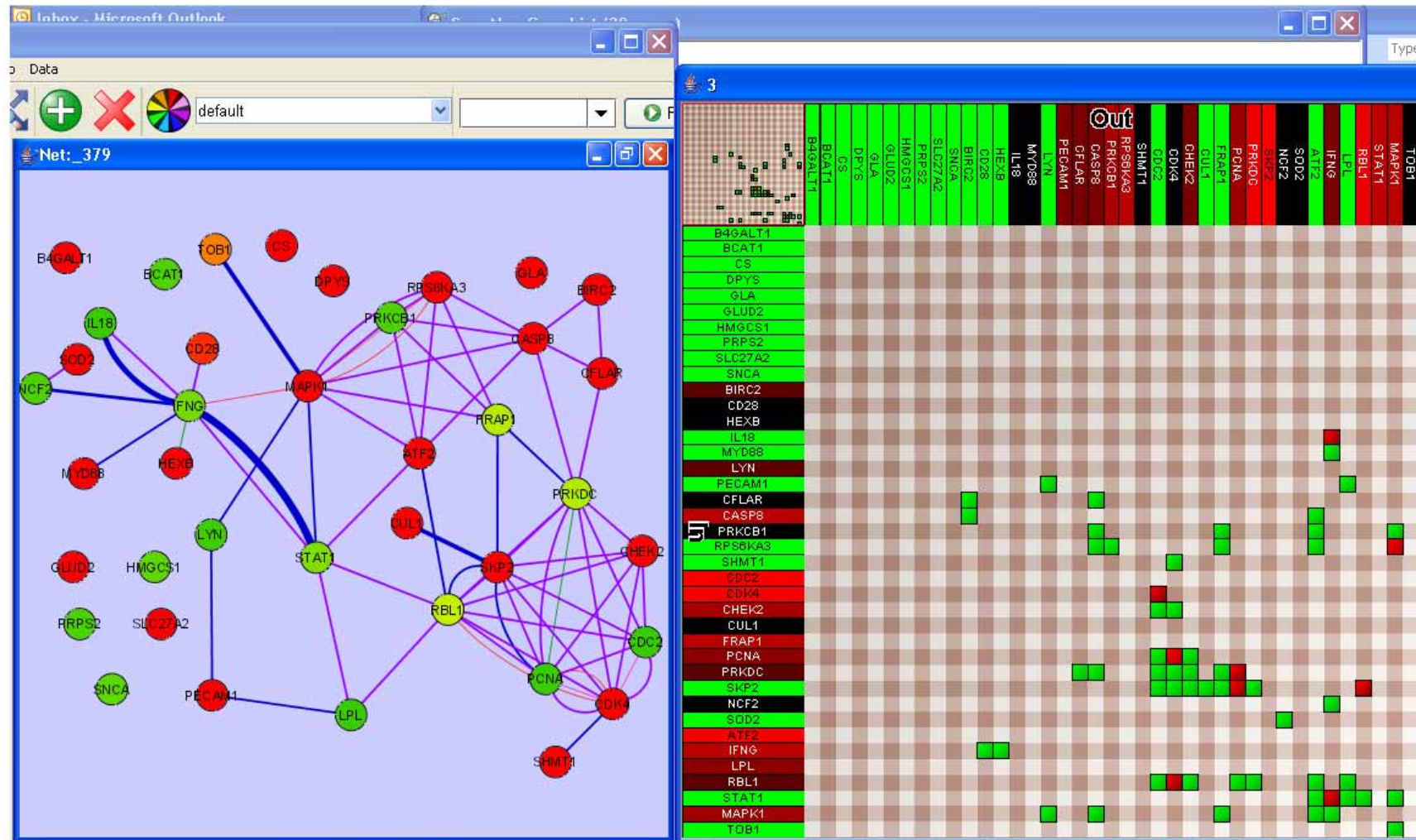
Simple graph, symmetric, binary



Directed and weighted



For more information: Diestel, R. (2010) *Graph Theory, 4th Edition*. Berlin, Heidelberg, Springer.

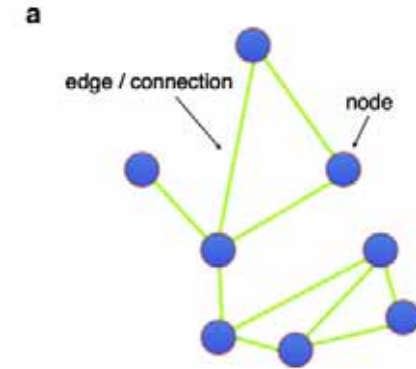


Jean-Daniel Fekete http://wiki.cytoscape.org/InfoVis_Toolkit

Fekete, J.-D. The infovis toolkit. Information Visualization, INFOVIS 2004, 2004. IEEE, 167-174.

Order = total number of nodes n ; **Size** = total number of links (a):

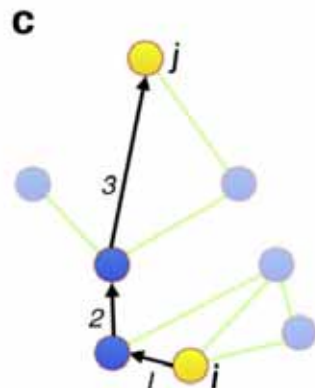
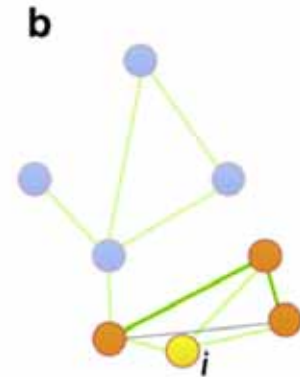
$$\sum_i \sum_j a_{ij}$$



Clustering Coefficient (b) = the degree of concentration of the connections of the node's neighbors in a graph and gives a measure of local inhomogeneity of the link density:

$$C_i = \frac{2t_i}{k(k_i - 1)}$$

$$C = \frac{1}{n} \sum_i C_i$$

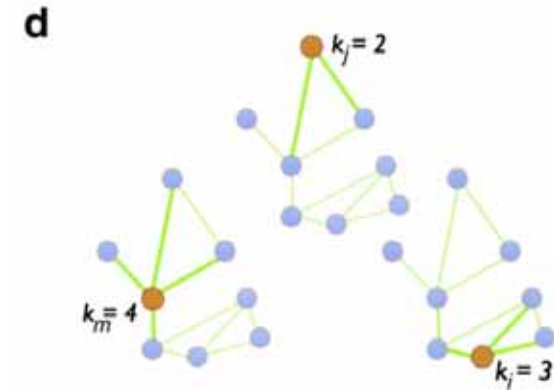


Path length (c) = is the arithmetical mean of all the distances:

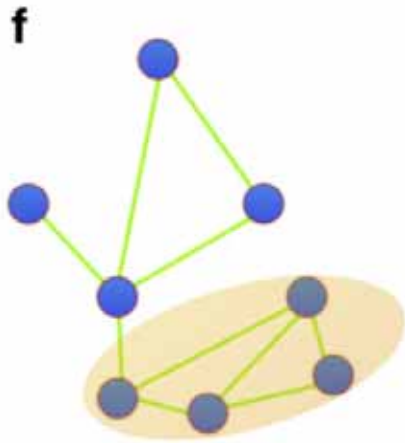
$$l = \frac{1}{n(n-1)} \sum_{i \neq j} d_{ij}$$

Costa, L. F., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. (2007) Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56, 1, 167-242.

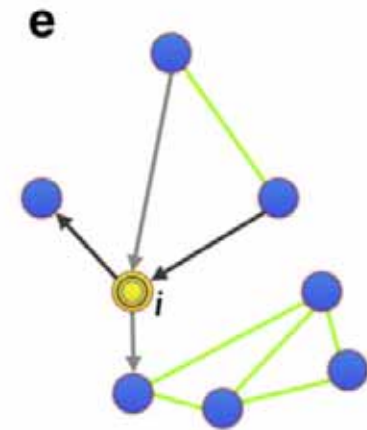
- Centrality (d) = the level of “betweenness- centrality” of a node i (“hub-node in Slide 28);



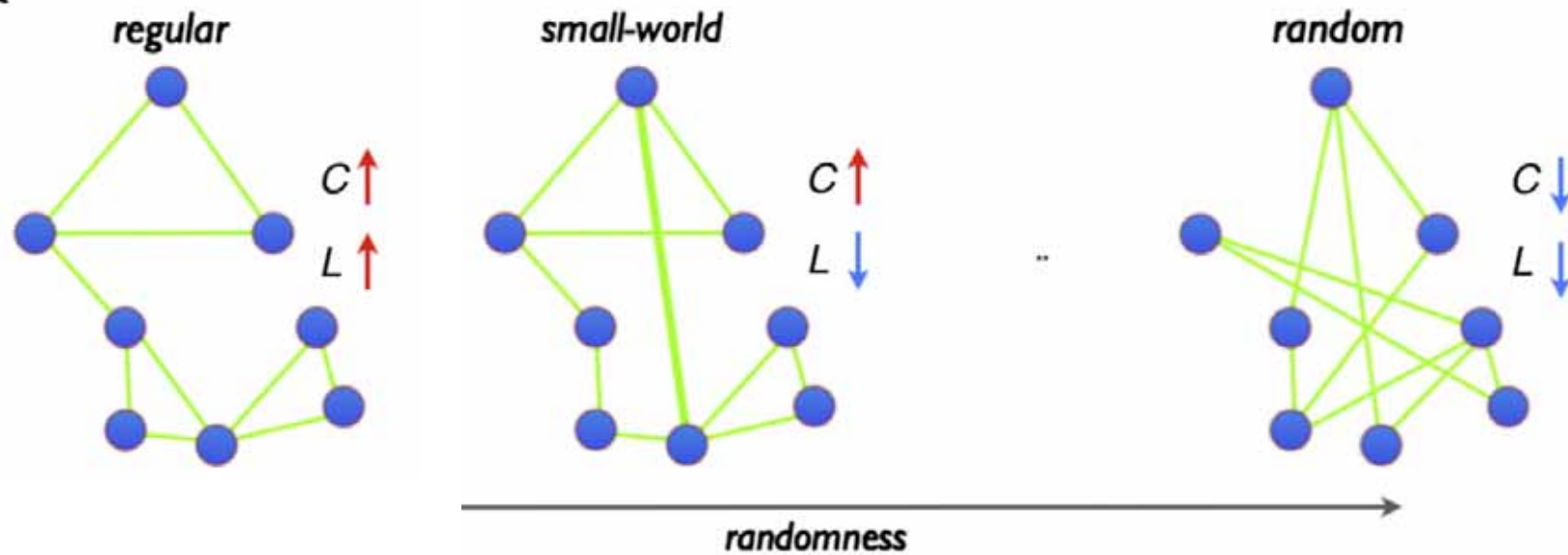
- Nodal degree (e) = number of links connecting i to its neighbors: $k_i = \sum_j a_{ij}$



Modularity (f) = describes the possible formation of communities in the network, indicating how strong groups of nodes form relative isolated sub-networks within the full network (refer also to Slide 5-8).

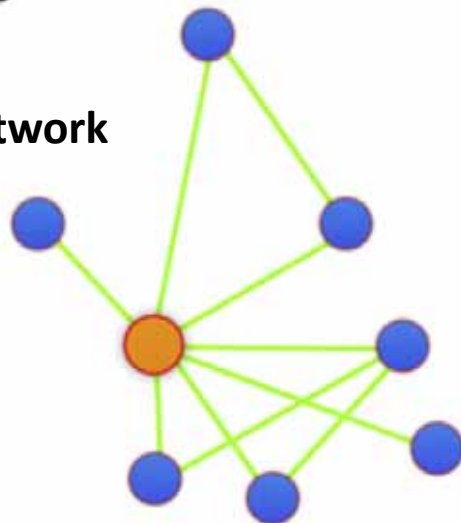


a

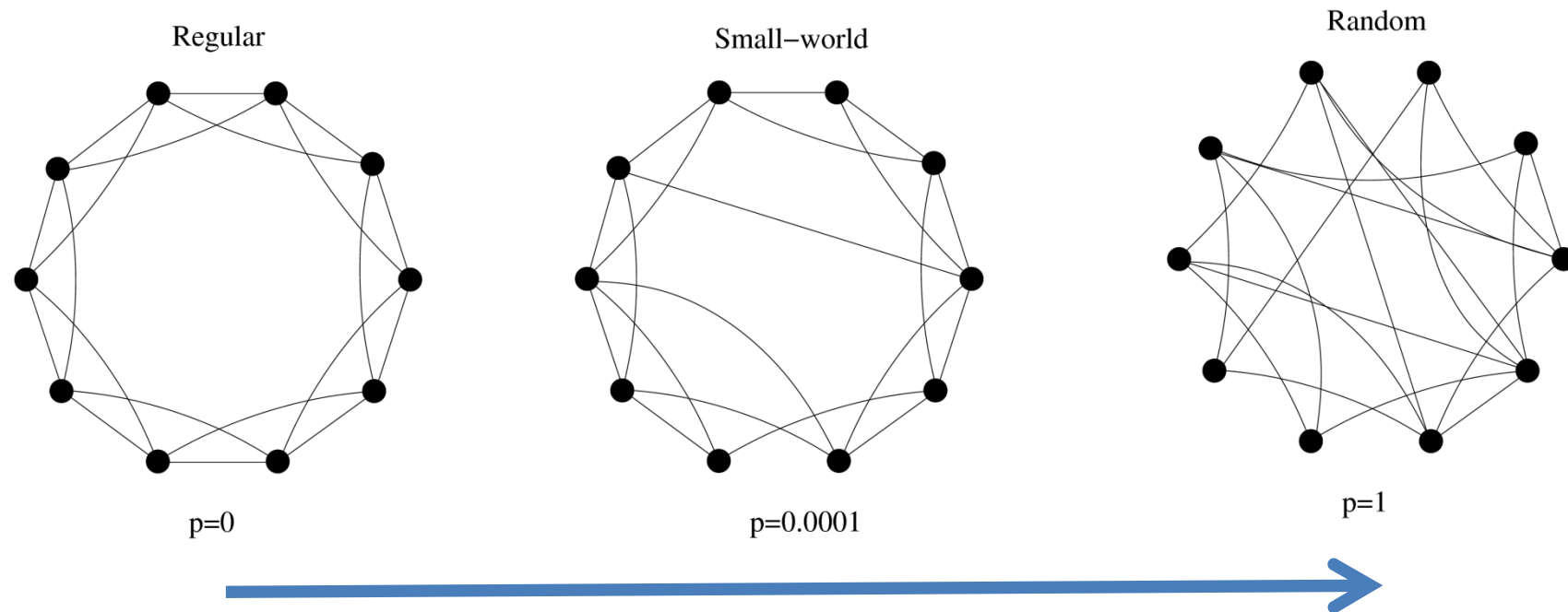


b

Scale-free network



Van Heuvel & Hulshoff (2010)

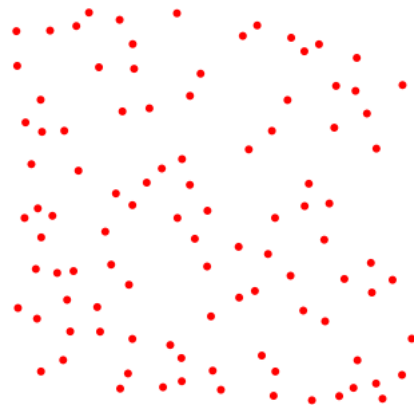


Increasing randomness

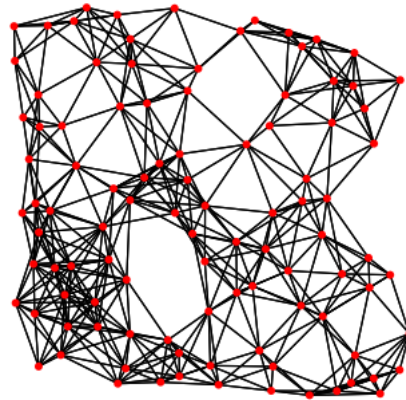
29.000 citations ...

Watts, D. J. & Strogatz, S. (1998) Collective dynamics of small-world networks. *Nature*, 393, 6684, 440-442.

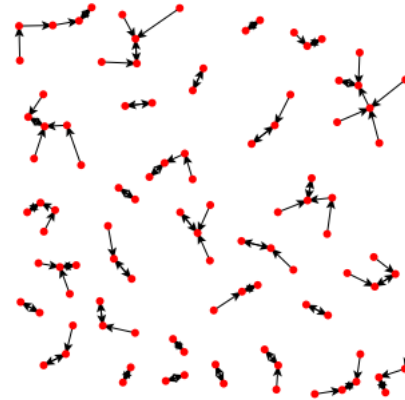
Milgram, S. 1967. The small world problem. *Psychology today*, 2, (1), 60-67.



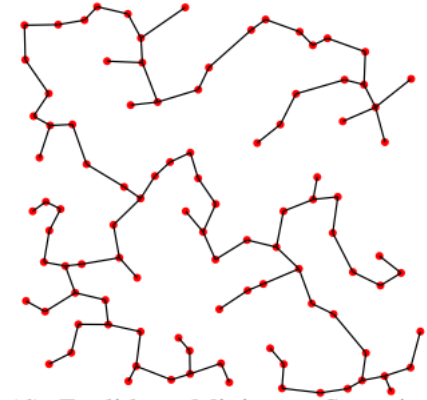
(a) Initial set of points.



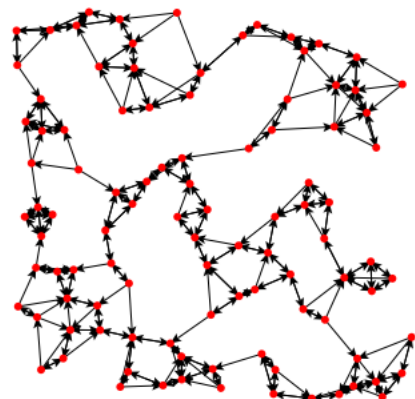
(b) 1-ball Graph.



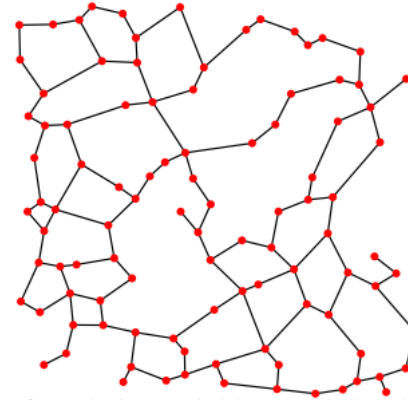
(c) 1-Nearest-Neighbor Graph.



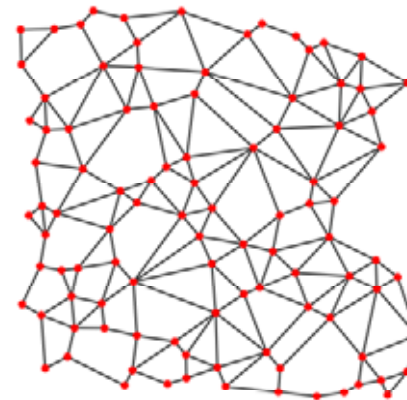
(d) Euclidean Minimum Spanning Tree.



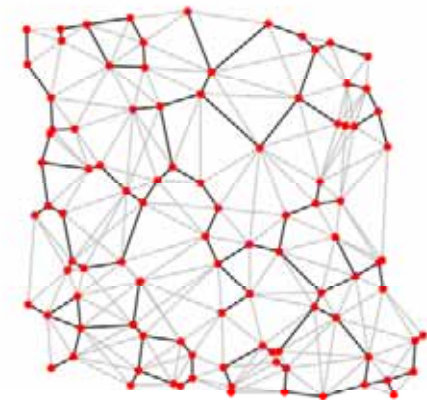
(e) 3-Nearest-Neighbor Graph.



(f) Relative Neighborhood Graph.

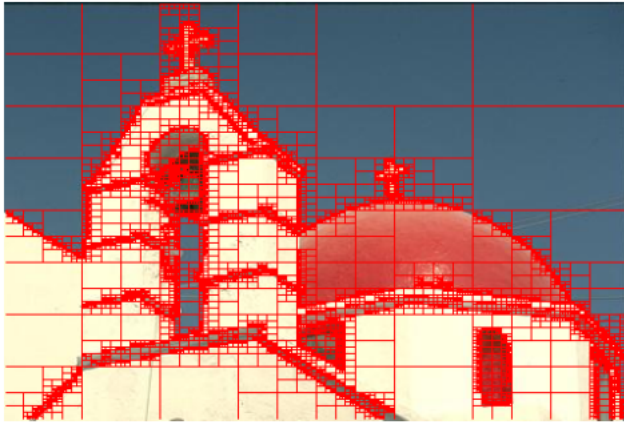


(g) Gabriel Graph.

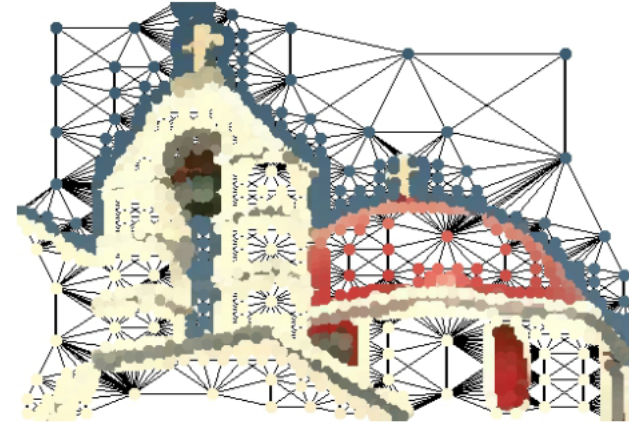
(h) β -Skeleton Graph, $\beta = 1.1$: black edges, $\beta = 0.9$: grey edges.

Lézoray, O. & Grady, L. 2012. Graph theory concepts and definitions used in image processing and analysis. In: Lézoray, O. & Grady, L. (eds.) *Image Processing and Analysing With Graphs: Theory and Practice*. Boca Raton (FL): CRC Press, pp. 1-24.

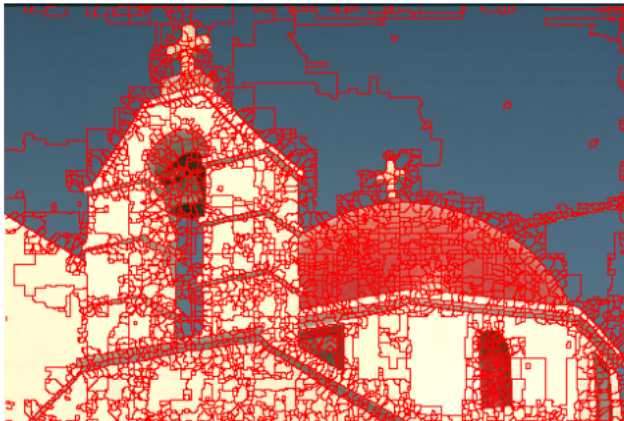
07 How do you get point cloud data from natural images?



a) quadtree tessellation



b) RAG assoc. to the quadtree



c) Watershed Algorithm



d) SLIC superpixels

Lézoray, O. & Grady, L. 2012. Graph theory concepts and definitions used in image processing and analysis. In: Lézoray, O. & Grady, L. (eds.) *Image Processing and Analysing With Graphs: Theory and Practice*. Boca Raton (FL): CRC Press, pp. 1-24.

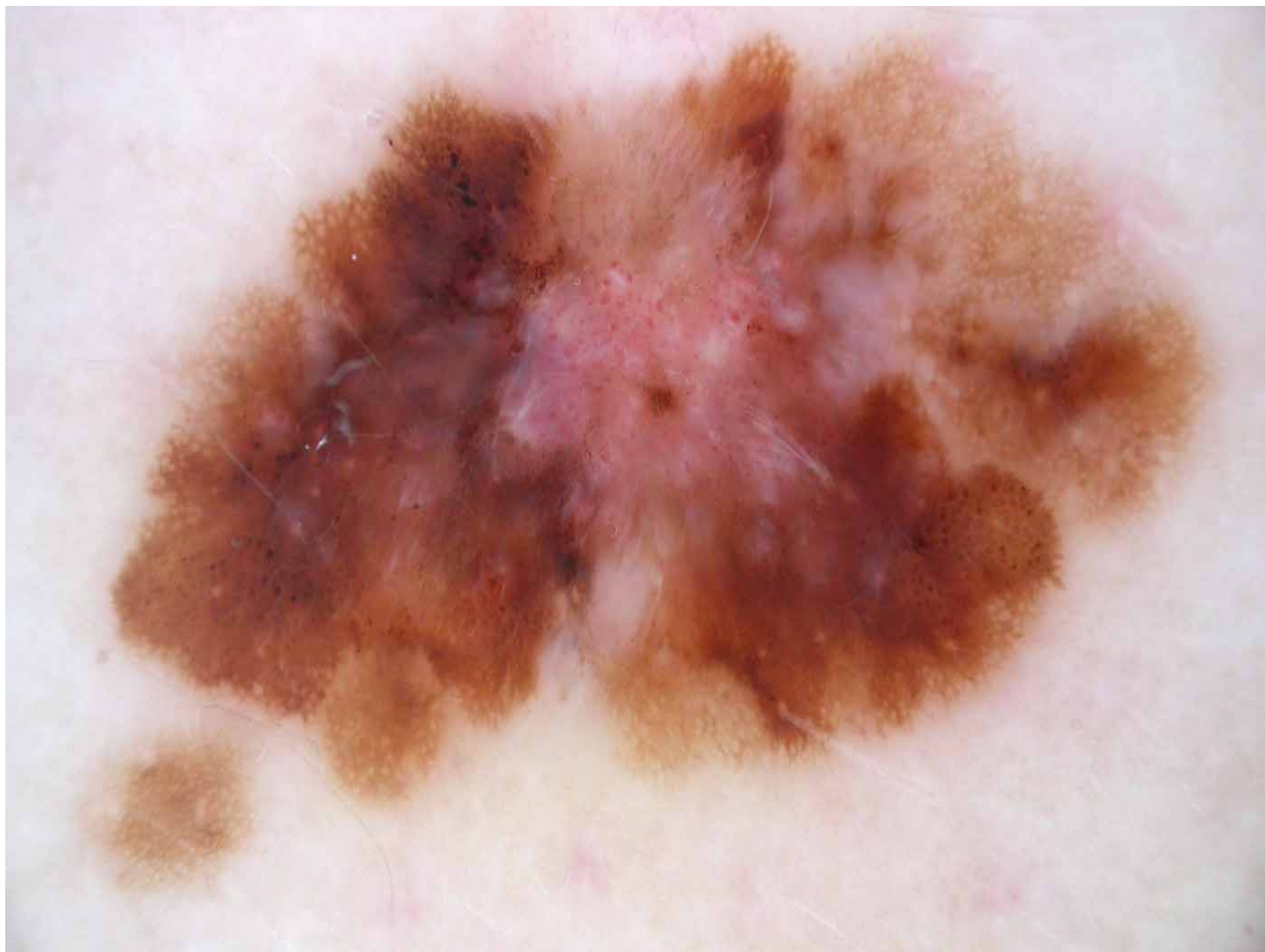
Algorithm 4.2 Watershed transform w.r.t. topographical distance based on image integration via the Dijkstra-Moore shortest paths algorithm.

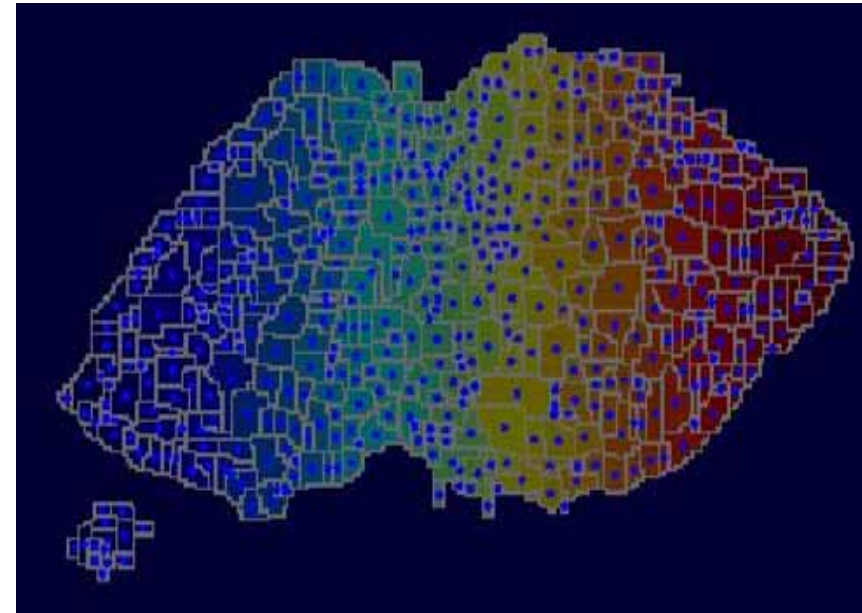
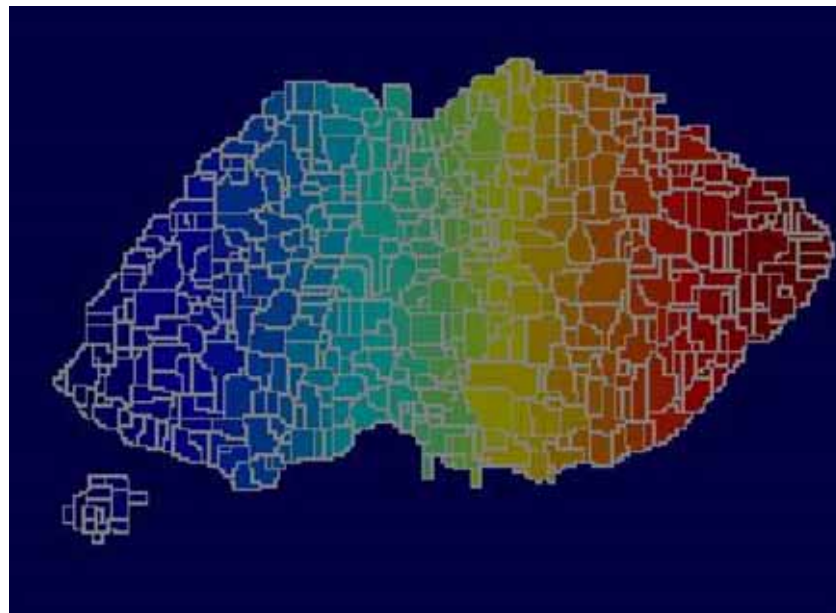
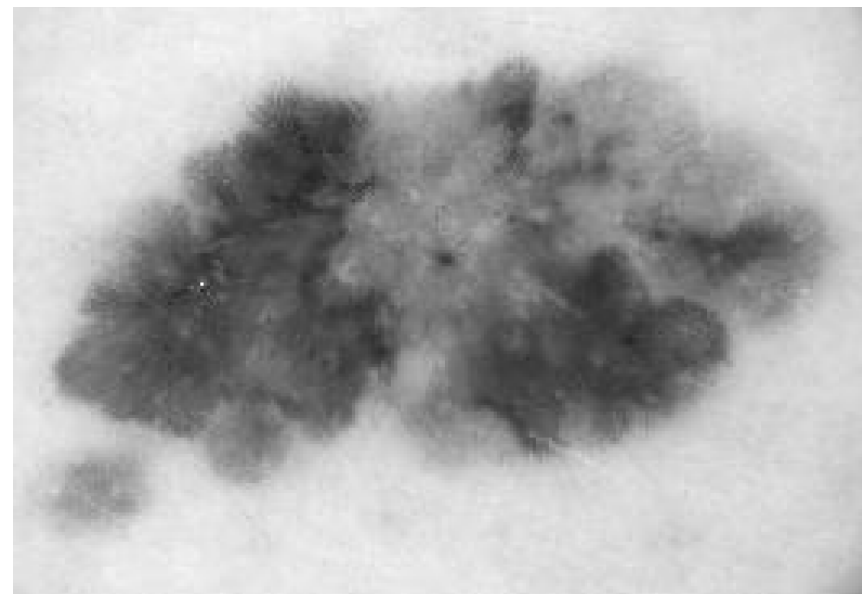
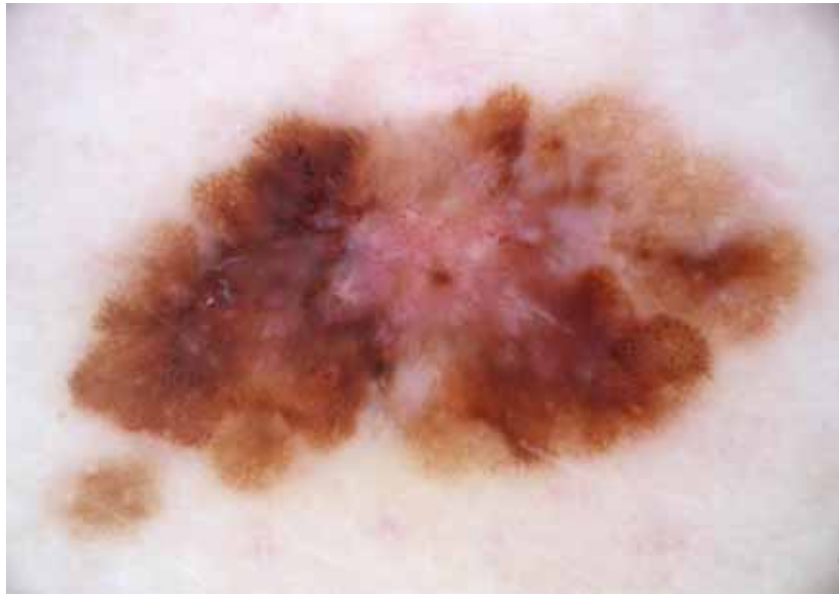
```

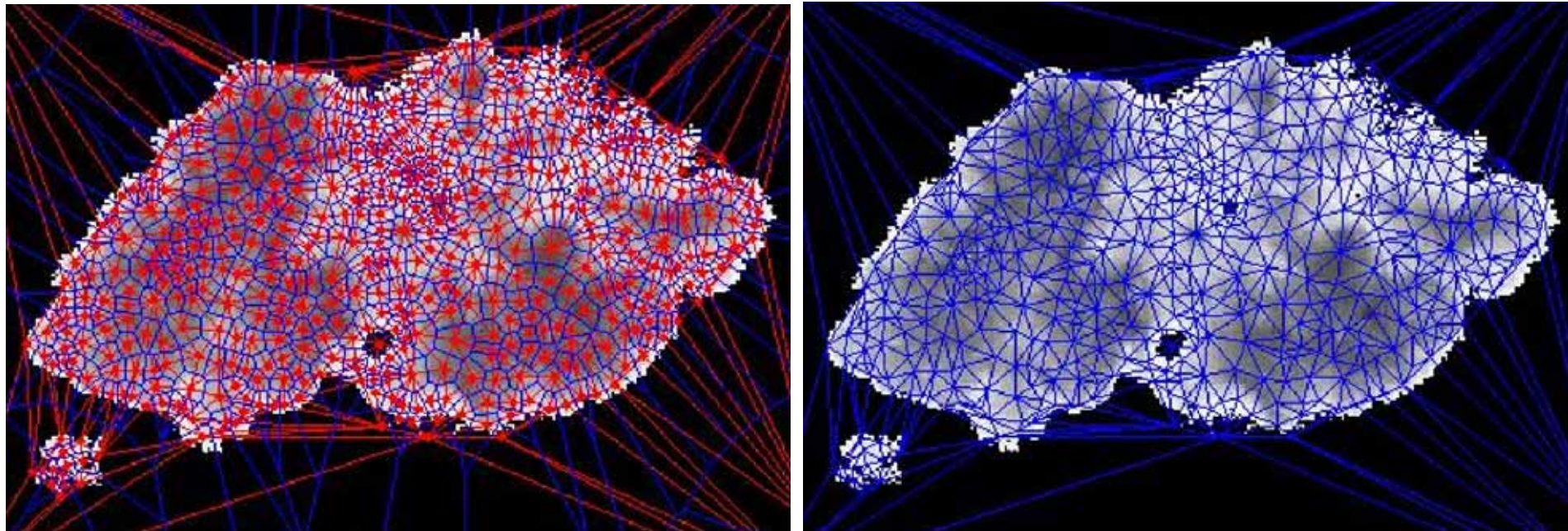
1: procedure ShortestPathWatershed;
2: INPUT: lower complete digital grey scale image  $G = (V, E, im)$  with cost function  $cost$ .
3: OUTPUT: labelled image  $lab$  on  $V$ .
4: #define WSHED 0 (* label of the watershed pixels *)
5: (* Uses distance image  $dist$ . On output,  $dist[v] = im[v]$ , for all  $v \in V$ . *)
6:
7: for all  $v \in V$  do (* Initialize *)
8:    $lab[v] \leftarrow 0$  ;  $dist[v] \leftarrow \infty$ 
9: end for
10: for all local minima  $m_i$  do
11:   for all  $v \in m_i$  do
12:      $lab[v] \leftarrow i$  ;  $dist[v] \leftarrow im[v]$  (* initialize distance with values of minima *)
13:   end for
14: end for
15: while  $V \neq \emptyset$  do
16:    $u \leftarrow GetMinDist(V)$  (* find  $u \in V$  with smallest distance value  $dist[u]$  *)
17:    $V \leftarrow V \setminus \{u\}$ 
18:   for all  $v \in V$  with  $(u, v) \in E$  do
19:     if  $dist[u] + cost[u, v] < dist[v]$  then
20:        $dist[v] \leftarrow dist[u] + cost(u, v)$ 
21:        $lab[v] \leftarrow lab[u]$ 
22:     else if  $lab[v] \neq WSHED$  and  $dist[u] + cost[u, v] = dist[v]$  and  $lab[v] \neq lab[u]$  then
23:        $lab[v] = WSHED$ 
24:     end if
25:   end for
26: end while

```

Meijster, A. & Roerdink, J. B. A proposal for the implementation of a parallel watershed algorithm. Computer Analysis of Images and Patterns, 1995. Springer, 790-795.







Holzinger, A., Malle, B. & Giuliani, N. 2014. On Graph Extraction from Image Data. In: Slezak, D., Peters, J. F., Tan, A.-H. & Schwabe, L. (eds.) Brain Informatics and Health, BIH 2014, Lecture Notes in Artificial Intelligence, LNAI 8609. Heidelberg, Berlin: Springer, pp. 552-563.

For Voronoi please refer to: Aurenhammer, F. 1991. Voronoi Diagrams - A Survey of a fundamental geometric data structure. *Computing Surveys*, 23, (3), 345-405.

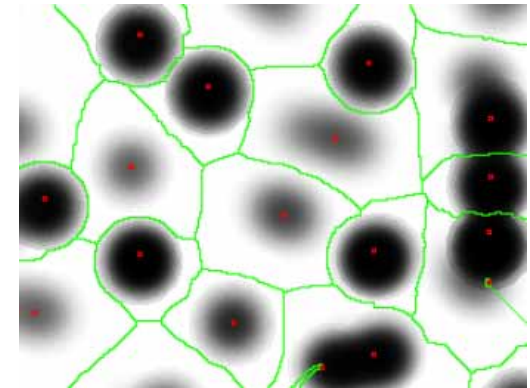
For Delaunay please refer to: Lee, D.-T. & Schachter, B. J. 1980. Two algorithms for constructing a Delaunay triangulation. *Intl. Journal of Computer & Information Sciences*, 9, (3), 219-242.

- More expressive data structures
- Find novel connections between data objects
- Fit for applying graph based machine learning techniques
- New approaches (Belief Propagation, global understanding from local properties)

Bunke, H.: Graph-based tools for data mining and machine learning. In Perner, P., Rosenfeld, A., eds.: Machine Learning and Data Mining in Pattern Recognition, Proceedings. Volume 2734 of Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin, (Berlin) 7–19

Holzinger, A., Blanchard, D., Bloice, M., Holzinger, K., Palade, V., Rabadan, R.: Darwin, Lamarck, or Baldwin: Applying evolutionary algorithms to machine learning techniques. In: The 2014 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2014), IEEE (2014) in print

- Topographic maps => landscapes with height structures
- Segmentation into regions of pixels
- Assuming drops of water raining on the map
- Following paths of descent
- Lakes called catchment basins
- Also possible: Flooding based
- Needs Topographical distance measures (MST)



Vincent, L. & Soille, P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE transactions on pattern analysis and machine intelligence, 13, (6), 583-598.

- 1) Transformation into a topographic map
 - Convert gray values into height information
- 2) Finding local minima
 - Inspecting small regions in sequence
- 3) Finding catchment basins
 - Algorithm simulating flooding
 - Graph algorithms such as Minimum Spanning Trees
- 4) Erecting watersheds
 - Artificial divide between catchment basins
 - Final segmentation lines

7	4	8	12	11	3
7	7	8	12	11	7
13	13	15	16	16	13
19	19	18	17	15	7
20	18	17	16	15	5

(a) The original image

→	m	←	←	→	m
↗	↑	↖	←	↗	↑
↑	↑	↖	↖	↗	↑
↑	↑	↑	→	↘	↓
→	→	→	→	→	m

(b) Each pixel connect to lowest minimum

0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	2	2	2
2	2	2	2	2	2

(c) The Image with labels

Connects each pixel to the lowest neighbor pixel, all pixel connected to same lowest neighbor pixel form a segment

- Region Merging
 - Based on Kruskals MST algorithm
 - Takes input image as natural graph with vertices := pixels and edges := pixel neighborhoods
 - Visits edges in ascending order of weight and merges regions if they satisfy a certain criterion
 - Flexible as merging criterion can be adapted as desired (for amount, size, or shape of resulting regions)

Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision 59 (2004) 167–181

- We want to find “interesting” novel patterns (rules, anomalies, outliers, similarities, ...)
- Problem #1: How to get a graph?
- Problem #2: How do graphs evolve?
- Problem #3: What tools to apply?
- Problem #4: Scalability to TB, PB, EB ...
- **Success is in repeatability and scalability**

- Study of complex networks started in the 1990s with the insight that real networks contain properties not present in random (Erdős-Renyi) networks.
- Meanwhile networks and network-based approaches form an integral part of many studies throughout the sciences.
- Graph-Theory provides powerful tools to organize data structurally and in combination with statistical and machine learning methods allows a meaningful analysis of underlying processes.
- For instance, a mapping of causal disease genes and disorders as made available by the OMIM database provided novel insights into disease patterns, as recently demonstrated by investigating the diseasome (<http://diseasome.eu>).



Thank you!

Questions

- 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?

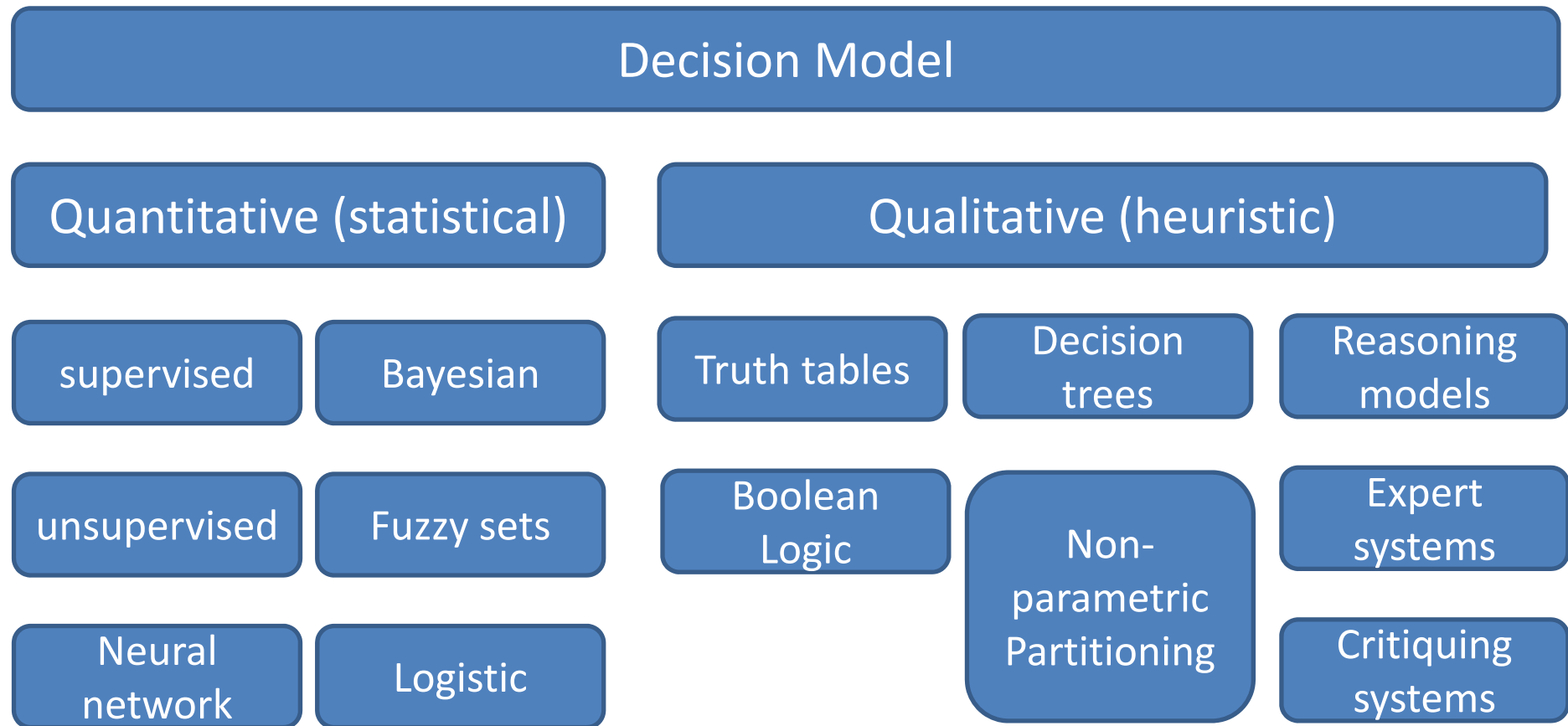
- Why do we want to apply ML to graphs?
- Describe typical ML tasks on the example of blood cancer cells!
- If you have a set of points – which similarity measures are useful?
- Why is graph comparison in the medical domain useful?
- Why is the Gromov-Hausdorff distance useful?
- What is the central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

- 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 trial – 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?

- 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 $\text{imply}(A,B)$, 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.

Appendix

1) Reasoning under Uncertainty



Bemmel, J. H. v. & Musen, M. A. (1997) *Handbook of Medical Informatics*. Heidelberg, Springer.

- The information available to humans is often imperfect – imprecise - uncertain.
- This is especially in the medical domain the case.
- An **human agent** can cope with deficiencies.
- Classical logic permits only **exact reasoning**:
- IF A is true THEN A is non-false and
IF B is false THEN B is non-true
- Most real-world problems do not provide this exact information, mostly it is inexact, incomplete, uncertain and/or **un-measurable!**

- MYCIN is a rule-based Expert System, which is used for therapy planning for patients with bacterial infections
- Goal oriented strategy (“Rückwärtsverkettung”)
- To every rule and every entry a certainty factor (CF) is assigned, which is between 0 und 1
- Two measures are derived:
 - MB: measure of belief
 - MD: measure of disbelief
- Certainty factor – CF of an element is calculated by:
$$CF[h] = MB[h] - MD[h]$$
- CF is positive, if more evidence is given for a hypothesis, otherwise CF is negative
- $CF[h] = +1 \rightarrow h$ is 100 % true
- $CF[h] = -1 \rightarrow h$ is 100% false

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

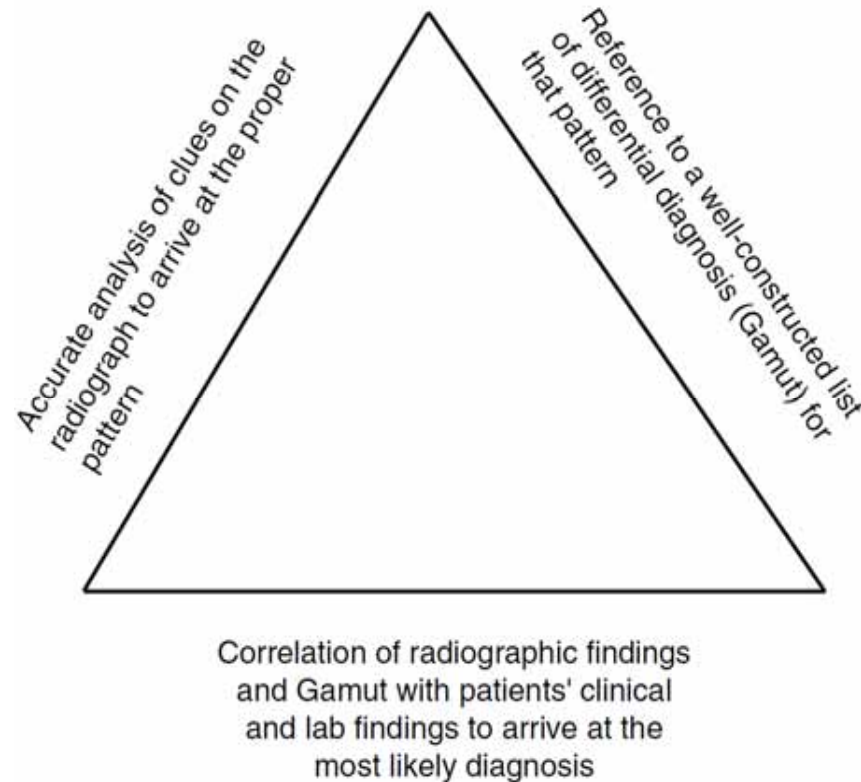
$CF[h_2, E] = -.3$: There is weakly suggestive evidence (.3) that PATIENT-1 is not febrile

$CF[h_3, E] = +1$: It is definite (1) that the name of PATIENT-1 is John Jones

Shortliffe, E. H. & Buchanan, B. G. (1984) *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley.

<https://www.youtube.com/watch?v=IVGWM0CKNWA> (“real nurse triage”)





Reeder, M. M. & Felson, B. 2003.
Reeder and Felson's gamuts in radiology: comprehensive lists of roentgen differential diagnosis, New York, Springer Verlag.

Gamut F-137

PHRENIC NERVE PARALYSIS OR DYSFUNCTION

COMMON

1. Iatrogenic (eg, surgical injury; chest tube; therapeutic avulsion or injection; subclavian vein puncture)
2. Infection (eg, tuberculosis; fungus disease; abscess)
3. Neoplastic invasion or compression (esp. carcinoma of lung)

UNCOMMON

1. Aneurysm_g, aortic or other
2. Birth trauma (Erb's palsy)
3. Herpes zoster
4. Neuritis, peripheral (eg, diabetic neuropathy)
5. Neurologic disease_g (eg, hemiplegia; encephalitis; polio; Guillain-Barré S.)
6. Pneumonia
7. Trauma

Reference

1. Prasad S, Athreya BH: Transient paralysis of the phrenic nerve associated with head injury. JAMA 1976;236:2532-2533

REEDER AND FELSON'S

GAMUTS IN RADIOLOGY

GAMUT G-25EROSIVE GASTRITIS*

COMMON

1. Acute gastritis (eg, alcohol abuse)
2. Crohn's disease **I** **I**
3. Drugs (eg, aspirin **I** **I**; NSAID **I**; steroids)
4. *Helicobacter pylori* infection **I**
5. Idiopathic
6. [Normal areae gastricae **I**]
7. Peptic ulcer; hyperacidity

UNCOMMON

1. Corrosive gastritis **I**
2. *Cryptosporidium* antritis
3. [Lymphoma]
4. Opportunistic infection (eg, candidiasis {moniliasis} **I**; herpes simplex; cytomegalovirus)
5. Postoperative gastritis
6. Radiation therapy
7. Zollinger-Ellison S. **I**; multiple endocrine neoplasia (MEN) S.

* Superficial erosions or aphthoid ulcerations seen especially with double contrast technique.

[] This condition does not actually cause the gamuted imaging finding, but can produce imaging changes that simulate it.

Reeder, M. M. & Felson, B. (2003) *Reeder and Felson's gamuts in radiology: comprehensive lists of roentgen differential diagnosis*. New York, Springer Verlag.

<http://rfs.acr.org/gamuts/data/G-25.htm>

- Take patient information, e.g., observations, symptoms, test results, -omics data, etc. etc.
- Reach conclusions, and predict into the future, e.g. how likely will the patient be re-admissioned
- Prior = belief before making a particular observation
- Posterior = belief after making the observation and is the prior for the next observation – intrinsically incremental


$$p(x_i|y_j) = \frac{p(y_j|x_i)p(x_i)}{\sum p(x_i, y_j)p(x_i)}$$

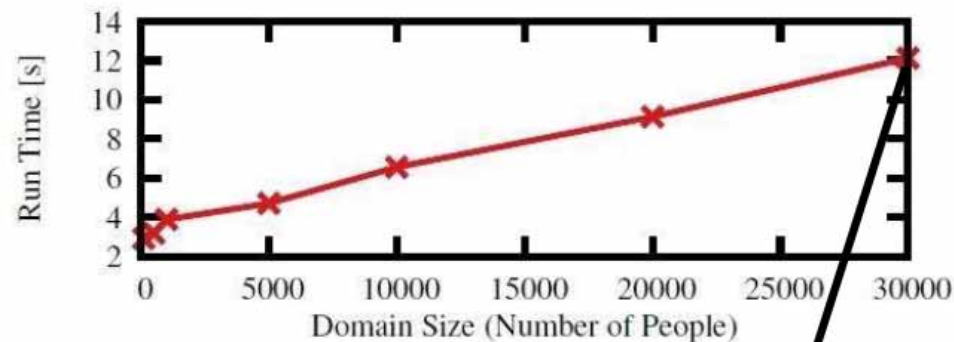
- **Type 1 Decisions:** related to the diagnosis, i.e. computers are used to assist in diagnosing a disease on the basis of the individual patient data. Questions include:
 - What is the probability that this patient has a myocardial infarction on the basis of given data (patient history, ECG, ...)?
 - What is the probability that this patient has acute appendices, given the signs and symptoms concerning abdominal pain?

- **Type 2 Decisions:** related to therapy, i.e. computers are used to select the best therapy on the basis of clinical evidence, e.g.:
 - What is the best therapy for patients of age x and risks y , if an obstruction of more than z % is seen in the left coronary artery?
 - What amount of insulin should be prescribed for a patient during the next 5 days, given the blood sugar levels and the amount of insulin taken during the recent weeks?

Bemmel, J. H. V. & Musen, M. A. 1997. *Handbook of Medical Informatics*, Heidelberg, Springer.

The future is in integrative ML, i.e. combining relational databases, ontologies and logic with probabilistic reasoning models and statistical learning – and algorithms that have good **scalability**

$\text{w Smokes}(x) \wedge \text{Friends}(x,y) \Rightarrow \text{Smokes}(y)$



~~Big data~~

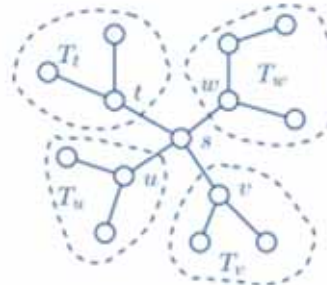
Big models

Learns a model over
900,030,000 random variables

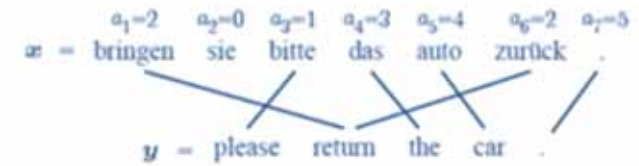
Van Den Broeck, G., Taghipour, N., Meert, W., Davis, J. & De Raedt, L. Lifted probabilistic inference by first-order knowledge compilation. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three, 2011. AAAI Press, 2178-2185.



1



2



3

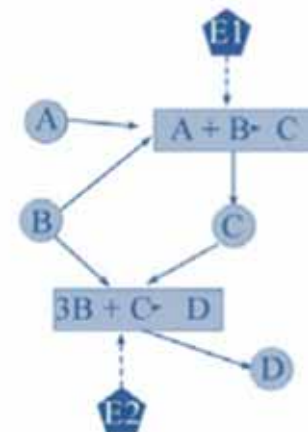
$$E(U \mid d) = \sum_{x_1, \dots, x_n} P(x_1, \dots, x_n \mid d) U(x_1, \dots, x_n, d)$$

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