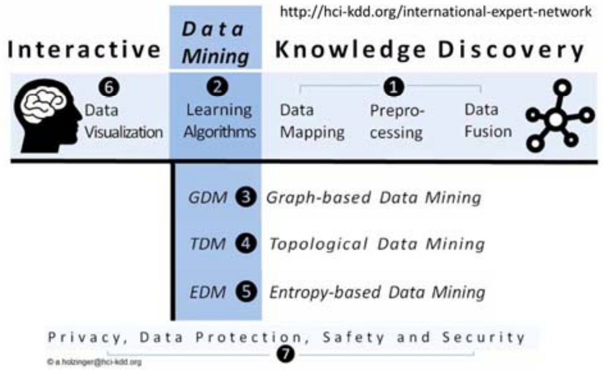
Science is to test crazy ideas – Engineering is to put these ideas into Business

Andreas Holzinger
VO 709.049 Medical Informatics
23.11.2016 11:15-12:45

Lecture 05 Probabilistic Graphical Models I: From Knowledge Representation to Graph Learning

a.holzinger@tugraz.at
Tutor: markus.plass@student.tugraz.at
http://hci-kdd.org/biomedical-informatics-big-data

ML needs a concerted effort fostering integrated research



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

Keywords

- Reasoning
- Uncertainty
- Graphs
- Complexity
- Graph structures
- Network Medicine
- Knowledge Spaces
- Biomedical Networks
- Emergence
- Robustness
- Modularity
- Structure Learning

Advance Organizer (1/3) A-G

- **Adjacency matrix** = simplest form of computational graph representation, in which 0 or 1 denotes whether or not there is a directed edge from one node to another (in graph theory adjacent nodes in a graph are linked by an edge);
- **Artifacts** = not only a noise disturbance, which is contaminating and influencing the signal (surrogates) but also data which is wrong, however interpreted as to be reliable, consequently may lead to a wrong decision;
- **Computational graph representation** = e.g. by adjacency matrices
- **Data fusion** = data integration techniques that analyze data from multiple sources in order to develop insights in ways that are more efficient and potentially more accurate than if they were developed by analyzing a single source of data. Signal processing techniques can be used to implement some types of data fusion (e.g. combined sensor data in Ambient Assisted Living);
- **Global Distance Test (GDT)** = a measure of similarity between two protein structures with identical amino acid sequences but different tertiary structures. It is most commonly used to compare the results of protein structure prediction to the experimentally determined structure as measured by X-ray crystallography or protein NMRM;
- **Graph theory** = study of mathematical structures to model relations between objects from a certain collection;
- **Graphs** = a hypothetical structure consisting of a series of nodes connected by weighted edges (graphs can be directed/undirected and stoichometric/non-stoichometric regarding interaction classes);

Advance Organizer (2/3) H-P

- **Homology** = in mathematics (especially algebraic topology and abstract algebra), it is (ὁμόιος homos = "identical") a certain general procedure to associate a sequence of Abelian groups (i.e. does not depend on their order) or modules with a given mathematical object such as a topological space or a group;
- **Homology modeling** = comparative modeling of protein, refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"); in Bioinformatics, homology modeling is a technique that can be used in molecular medicine.
- **In silico** = via computer simulation, in contrast to in vivo (within the living) or in vitro (within the glass);
- **Multi-scale representation** = in a graph, nodes do not have to represent biological objects on the same scale, one node (e.g. a molecule) may have an edge connecting it to a node representing a cell or tissue (the edge indicates that the molecule exerts an effect on the cell/tissue);
- **Network** = graphs containing cycles or alternative paths;
- **Network analysis** = a set of techniques used to characterize relationships among discrete nodes in a graph or a network;
- **Network topology** = the shape or structure of a network;
- **Petri-Net** = a special class of graph, consisting of two general classes or node: place and transition nodes;
- **Predictive modeling** = a set of techniques in which a mathematical model is created or chosen to best predict the probability of an outcome (e.g. regression);
- **P-System** = addresses the slowness of Petri-nets

Advance Organizer (3/3) R-V

- **Radius of a graph** = average minimum path length (biological networks are not arranged in a regular or symmetrical pattern);
- **Scale-free Topology** = ensures that there are very short paths between any given pair of nodes, allowing rapid communication between otherwise distant parts of the network (e.g. the Web has such a topology);
- **Semi-structured data** = does not conform with the formal structure of tables/data models assoc. with relational databases, but at least contains tags/markers to separate semantic elements and enforce hierarchies of records and fields within the data; aka schemaless or self-describing structure; the entities belonging to the same class may have different attributes even though they are grouped together;
- **Spatial analysis** = a set of techniques, applied from statistics, which analyze the topological, geometric, or geographic properties encoded in a data set;
- **Structural homology** = similar structure but different function;
- **Supervised learning** = machine learning techniques that infer a function or relationship from a set of training data (e.g. classification and support vector machines);
- **Time series analysis** = set of techniques from both statistics and signal processing for analyzing sequences of data points, representing values at successive times, to extract meaningful characteristics from the data;
- **Time series forecasting** = use of a model to predict future values of a time series based on known past values of the same or other series (e.g. structural modeling); decomposition of a series into trend, seasonal, and residual components, which can be useful for identifying cyclical patterns in the data;
- **Unstructured data** = complete randomness, noise; (wrongly, text is called unstructured, but there is some structure, too, so text data is a kind of weakly structured data);
- **Vertex degree** = within a topology, the numbers of edges connecting to a node;

Glossary

- ANSI = American National Standards Institute
- CD = cardiac development
- CDA = Clinical Document Architecture
- CHD = congenital heart disease
- CMM = Correlated motif mining
- DPI = Dossier Patient Integre´ = integrated patient record
- E = Edge
- EPR = Electronic Patient Record
- G(V,E) = Graph
- GI = gastrointestinal
- HER = Electronic Health Record
- HL7 = Health Level 7
- KEGG = Kyoto Encyclopedia of Genes and Genomes
- NP = nondeterministic polynomial time
- OWL = Web Ontology Language
- PPI = Protein-Protein Interaction
- SGML = Standard Generalized Markup Language
- TF= Transcription factor
- TG = Target Gene
- V = Vertex
- XML = Extensible Markup Language

Learning Goals … at the end of this lecture you …

- … have an idea of the **complexity of data** in biomedical informatics
- … are aware of the enormous importance of **graphs** (=network structures) and graph theory
- … have seen some application examples of **network structures** from both macro-cosmos and micro-cosmos and are fascinated about it;
- … have a rough overview about some basics of how to **get point clouds** out of data sets
- … have an understanding of some challenges of **network science**

Agenda for today

- **00 Reflection – follow-up from last lecture**
- **01 Reasoning under Uncertainty**
- **02 Where do graphs come from?**
- **03 Why are graphs so awesome?**
- **04 Knowledge Representation in Networks**
- **05 Graphs: Concepts, Metrics, Measures**
- **06 Example: Graphs from Natural Images**
- **07 Graphical Model Learning**

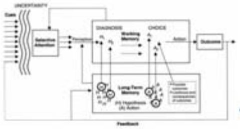
00 Reflection

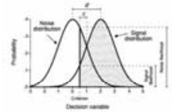
Reflection Quiz


1


2


3


4

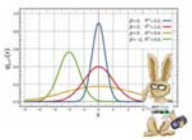

5


6

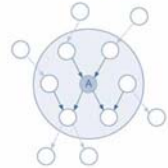
$$posterior = \frac{likelyhood * prior}{evidence}$$
7

$$h_{MAP} \equiv \underset{h \in H}{\operatorname{argmax}} P(h \mid D)$$
8

Preview-Quiz: Which concepts can you identify …



$$p(x) = \prod_i p(x_i | x_{pa_i})$$

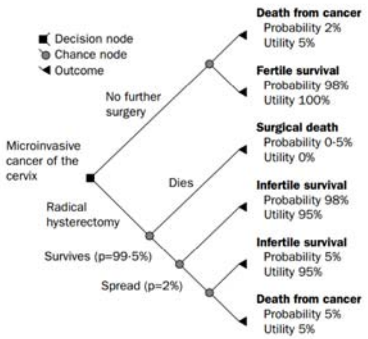


$$P(\mathbf{x}) = \prod_{i \in V} P_i(x_i) \prod_{(i,j) \in E} \frac{P_{ij}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

$$= P_1(x_1) P_{2|1}(x_2|x_1) P_{3|1}(x_3|x_1) P_{4|1}(x_4|x_1)$$

Graphical models are graphs where the **nodes** represent random variables and the **links** represent statistical dependencies between variables; This provides us with a tool for **reasoning under uncertainty**

Remember: Decision trees are coming from Clinical Practice



Physician treating a patient approx. 480 B.C. Beazley (1963), Attic Red-figured Vase-Painters, 813, 96. Department of Greek, Etruscan and Roman Antiquities, Sully, 1st floor, Campana Gallery, room 43 Louvre, Paris

Elwyn, G., Edwards, A., Eccles, M. & Rovner, D. 2001. Decision analysis in patient care. The Lancet, 358, (9281), 571-574.

Expected Utility Theory $E(U|d)$ Neumann-Morgenstern

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

The expected utility of decision $D = d$ is

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

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

An optimal single decision is the decision $D = dmax$ whose expected utility is maximal:

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

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

What are Probabilistic Graphical Models?

- PGM can be seen as a combination between
- **Graph Theory + Probability Theory + Machine Learning**
- One of the most exciting AI advances in the last decades
- Compact representation for exponentially-large probability distributions
- Example Question: "Is there a path connecting two proteins?"
- $Path\ (X,Y) := edge\ (X,Y)$
- $Path\ (X,Y) := edge\ (X,Y), path\ (Z,Y)$
- This can NOT be expressed in first-order logic
- Need a Turing-complete fully-fledged language

Friedman, N. 2004. Inferring cellular networks using probabilistic graphical models. Science, 303, (5659), 799-805.

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

01 Reasoning under Uncertainty

Key Challenges

- Medicine is an extremely complex application domain – dealing most of the time with uncertainties -> **probable information!**
- Key: Structure learning and prediction in large-scale biomedical networks with probabilistic graphical models
- Causal and Probabilistic Inference:
- Uncertainties are present at all levels in health related systems
- Data sets from which ML learns are noisy, mislabeled, atypical, etc. etc.
- Even with data of high quality, gauging and combining a multitude of data sources and constraints in usually imperfect models of the world requires us to represent and process uncertain knowledge in order to make viable decisions.
- In the increasingly complicated settings of modern science, model structure or causal relationships may not be known a-priori [1].
- Approximating probabilistic inference in Bayesian belief networks is NP-hard [2] -> here we need the "human-in-the-loop" [3]

[1] Sun, X., Janzing, D. & Schölkopf, B. Causal Inference by Choosing Graphs with Most Plausible Markov Kernels. ISAIM, 2006.
[2] Dagum, P. & Luby, M. 1993. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial intelligence, 60, (1), 141-153.
[3] Holzinger, A. 2016. Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Springer Brain Informatics (BRIN), 3, 1-13, doi:10.1007/s40708-016-0042-6.

Taxonomy of Decision Support Models

See lecture 8 for details!

Decision Model

Quantitative (statistical)

Qualitative (heuristic)

supervised | Bayesian | Truth tables | Decision trees | Reasoning models

unsupervised | Fuzzy sets | Boolean Logic | Non-parametric Partitioning | Expert systems

Neural network | Logistic | Critiquing systems

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

Dealing with uncertainty in the real world

- The information available to humans is often imperfect – imbalanced - 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!**

Original Example from MYCIN (See lecture 8 for details)

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

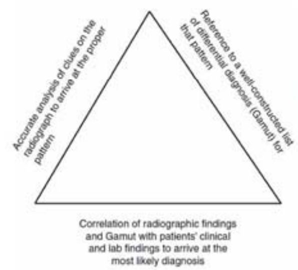
$CF[h_1,E] = .8$: There is strongly suggestive evidence (.8) that the identity of ORGANISM-1 is streptococcus

$CF[h_2,E] = -.3$: 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.*

Gamuts: Triangulation to find diagnoses



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, aortic or other
2. Birth trauma (Erb's palsy)
3. Herpes zoster
4. Neuritis, peripheral (eg, diabetic neuropathy)
5. Neurologic disease (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

Reasoning under uncertainty

- 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

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

Two types of decisions (Diagnosis vs. Therapy)

- **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.*

2) Where do Graphs come from?

Leonhard Euler (1707-1783) in 1736 …

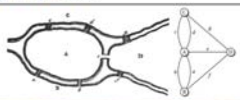
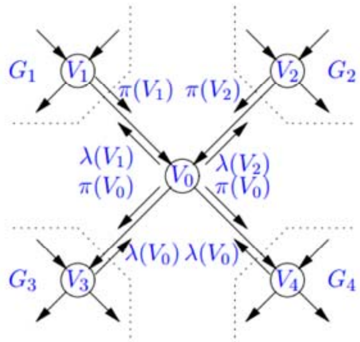


Image from https://people.kth.se/~carlofi/teaching/FEL3250-2013/courseinfo.html

252 years later: Belief propagation algorithm



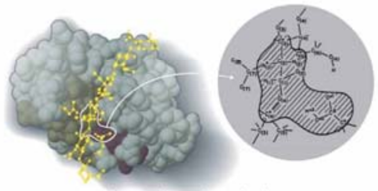
Pearl, J. 1988. Embracing causality in default reasoning. Artificial Intelligence, 35, (2), 259-271.

275 years later … the "Nobel-prize in Computer Science"

http://amturing.acm.org/vp/pearl_2658896.cfm

Nobel Prize in Chemistry 2013

http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013



http://news.harvard.edu/gazette/story/2013/10/nobel_prize_awarded_2013/

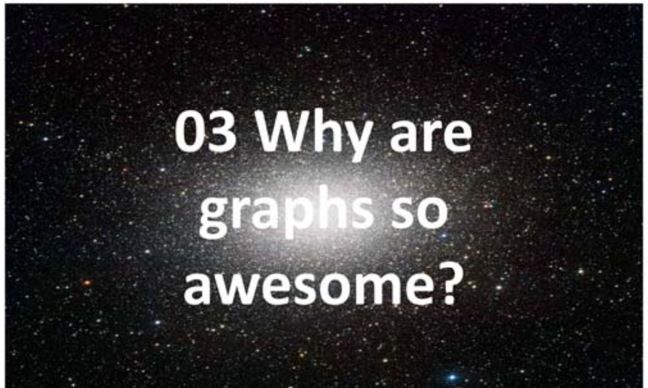
But our question was: Where do graphs come from?

- **Parametric models**
- given as direct input (**point cloud** data sets)
- Given as properties of a structure, e.g. biological networks
- Given as a representation of information (e.g. Facebook data, viral marketing, etc., …)

- **Nonparametric models**
- We extract the graph from other data [1]
- we learn the structure from samples and infer
- flat vector data, e.g. similarity graphs
- encoding structural properties (e.g. smoothness, independence, …)

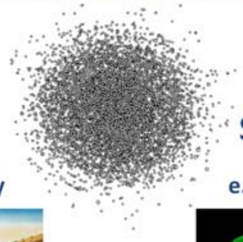
[1] Holzinger, A., Malle, B. & Giuliani, N. 2014. On Graph Extraction from Image Data. In: Slezak, D., Peters, J. F., Tan, A.-H. & Schwabe, L. (eds.) Brain Informatics and Health, BIH 2014, Lecture Notes in Artificial Intelligence, LNAI 8609. Heidelberg, Berlin: Springer, pp. 552-563, doi:10.1007/978-3-319-09891-3_50.

Our World in Data (1/2) – Macroscopic Structures



NGC 5139 Omega Centauri by Edmund Halley in 1677, ESO, Atacama, Chile

Two thematic mainstreams in dealing with data …



Time

e.g. Entropy

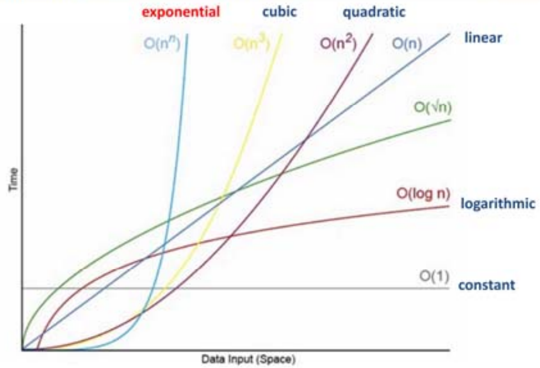
Space

e.g. Topology

Dali, S. (1931) The persistence of memory

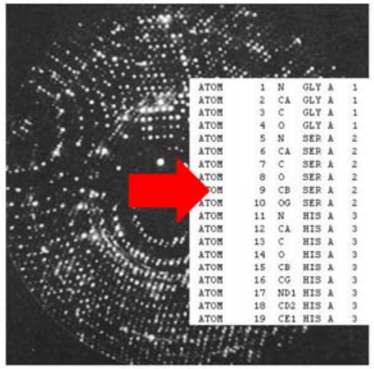
Bagula & Bourke (2012) Klein-Bottle

Complexity Problem: Time versus Space



P versus NP and the Computational Complexity Zoo, please have a look at https://www.youtube.com/watch?v=YX40hbAHx3s

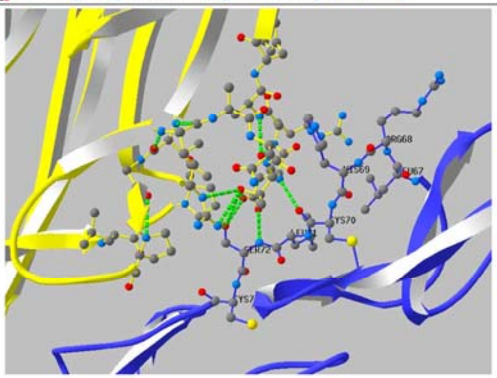
Our World in Data – Microscopic Structures



ATOM	1	N	GLY	A	1	44.842	51.034	101.284	0.01	27.20
ATOM	2	CA	GLY	A	1	45.640	50.230	100.389	0.01	26.99
ATOM	3	C	GLY	A	1	46.692	49.648	101.308	0.01	26.80
ATOM	4	O	GLY	A	1	46.895	50.222	102.381	0.01	26.91
ATOM	5	N	SER	A	2	47.283	48.516	100.951	1.00	26.26
ATOM	6	CA	SER	A	2	48.277	47.866	101.761	1.00	26.17
ATOM	7	C	SER	A	2	49.212	47.031	100.845	1.00	24.21
ATOM	8	O	SER	A	2	49.060	47.195	99.630	1.00	19.77
ATOM	9	CB	SER	A	2	47.438	47.091	102.800	1.00	26.31
ATOM	10	OG	SER	A	2	46.276	46.356	102.404	1.00	27.99
ATOM	11	N	HIS	A	3	50.147	46.186	101.370	1.00	23.93
ATOM	12	CA	HIS	A	3	51.129	45.389	100.609	1.00	21.44
ATOM	13	C	HIS	A	3	50.953	43.905	100.849	1.00	20.32
ATOM	14	O	HIS	A	3	50.530	43.595	101.950	1.00	22.00
ATOM	15	CB	HIS	A	3	52.555	45.674	100.990	1.00	19.69
ATOM	16	CG	HIS	A	3	52.940	47.090	100.611	1.00	21.44
ATOM	17	ND1	HIS	A	3	53.371	47.470	99.422	1.00	20.87
ATOM	18	CD2	HIS	A	3	52.956	48.175	101.433	1.00	21.69
ATOM	19	CE1	HIS	A	3	53.676	48.730	99.476	1.00	20.57

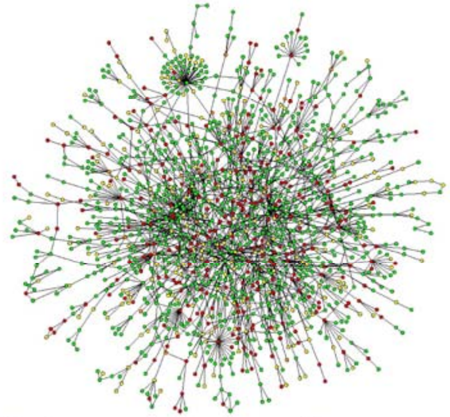
Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinformatics: Protein Structures with Physicochemical and Biological Annotations. In: *Central European Multimedia and Virtual Reality Conference. Prague, Czech Technical University (CTU), 69-74*

Getting Insight: Knowledge Discovery from Data



Wiltgen, M., Holzinger, A. & Tilz, G. P. (2007) Interactive Analysis and Visualization of Macromolecular Interfaces Between Proteins. In: *Lecture Notes in Computer Science (LNCS 4799). Berlin, Heidelberg, New York, Springer, 199-212.*

First yeast protein-protein interaction network

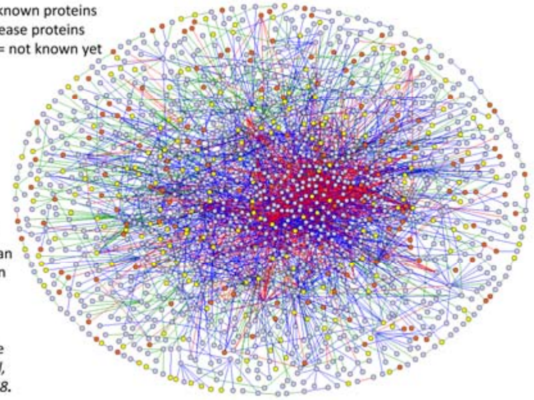


Nodes = proteins
Links = physical interactions (bindings)
Red Nodes = lethal
Green Nodes = non-lethal
Orange = slow growth
Yellow = not known

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

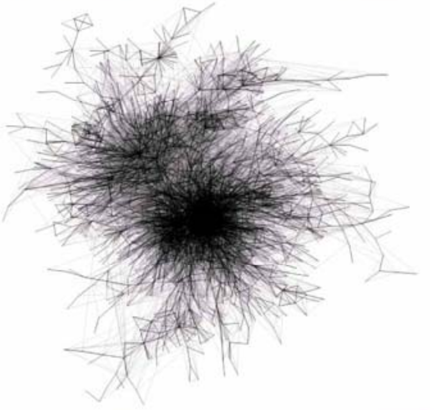
First human protein-protein interaction network

Light blue = known proteins
Orange = disease proteins
Yellow ones = not known yet



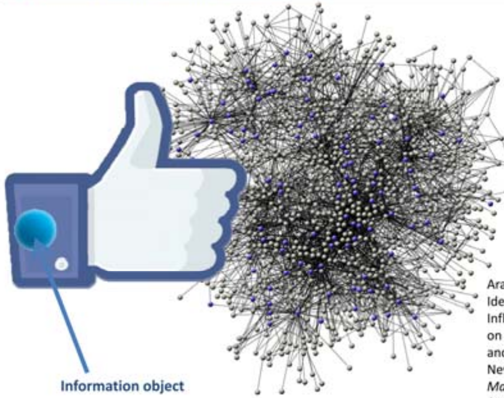
Stelzl, U. et al. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell, 122, 6, 957-968.*

Non-Natural Network Example: Blogosphere



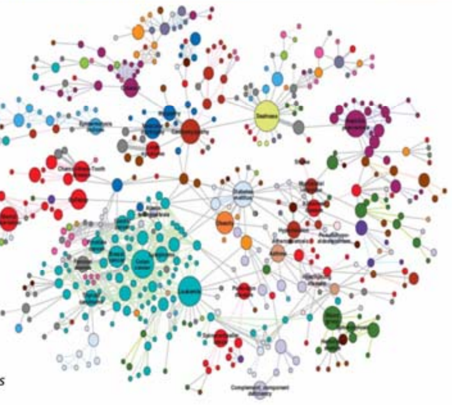
Hurst, M. (2007), Data Mining: Text Mining, Visualization and Social Media. Online available: http://datamining.typepad.com/data_mining/2007/01/the_blogosphere.html, last access: 2011-09-24

Social Behavior Contagion Network



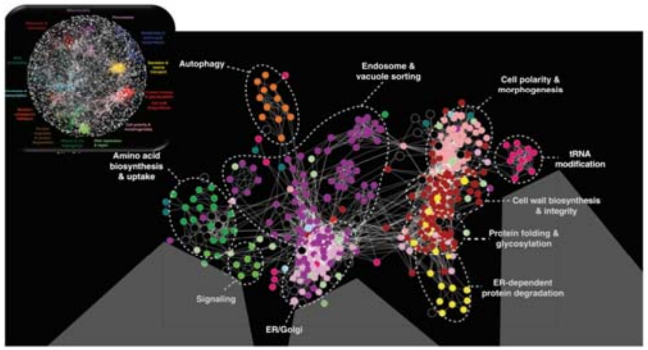
Aral, S. (2011) Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science, 30, 2, 217-223.*

Human Disease Network -> Network Medicine



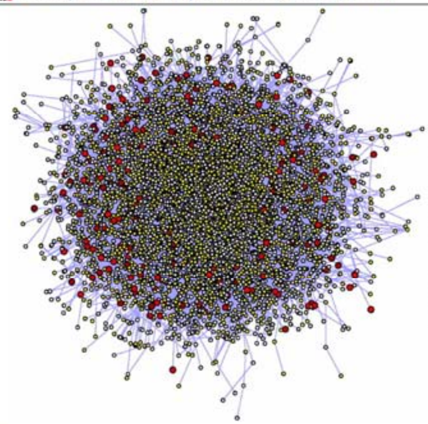
Barabási, A. L., Gulbahce, N. & Loscalzo, J. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, *12*, ***56-68.***

The Genetic Landscape of a cell



Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L., Toufighi, K. & Mostafavi, S. 2010. The genetic landscape of a cell. science, 327, (5964), 425-431.

Example for a weakly structured data set - PPI

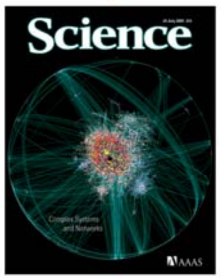


Kim, P. M., Korbel, J. O. & Gerstein, M. B. 2007. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. Proceedings of the National Academy of Sciences, 104, (51), 20274-20279.

04 Knowledge Representation in Network Medicine

Network Science – Graph Theory

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

Network of Networks in Biology

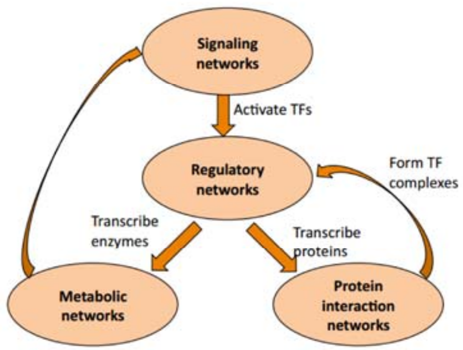
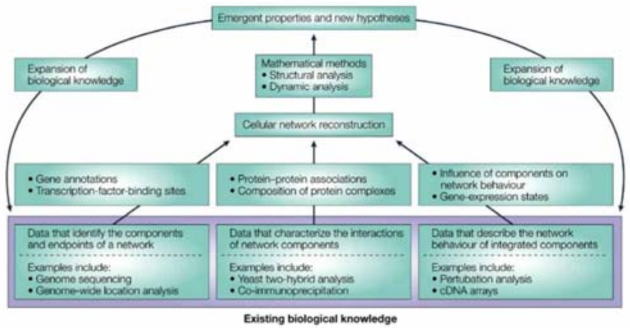


Image credit to Anna Goldenberg, Machine Learning Group, Toronto

From data sets to networks



Nature Reviews | Molecular Cell Biology

Image description find here:
http://www.nature.com/nrm/journal/v6/n2/fig_tab/nrm1570_F1.html

Regulatory>Metabolic>Signaling>Protein>Co-expression

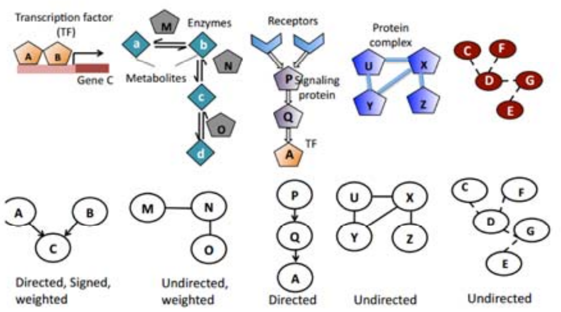
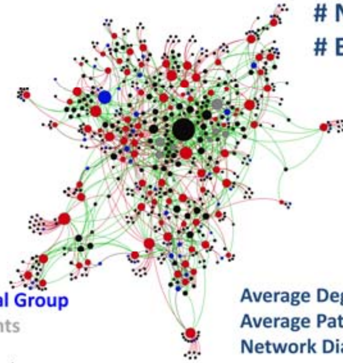


Image credit to Anna Goldenberg, Toronto

Example for a Medical Knowledge Space



Nodes: 641

Edges: 1250

Agent
Condition
Pharmacological Group
Other Documents

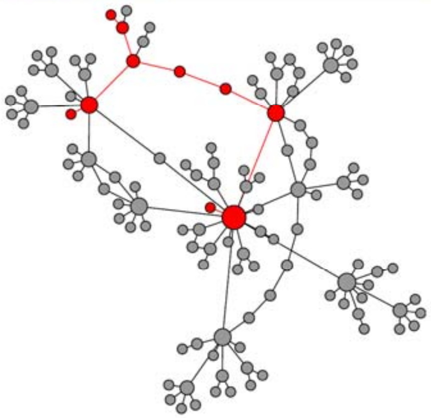
Average Degree: 3.888
Average Path Length: 4.683
Network Diameter: 9

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)

Medical Details of the Graph

- **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

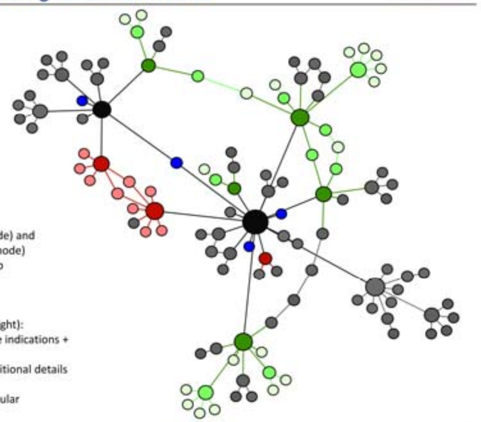
Example for the shortest path



Example for finding related structures

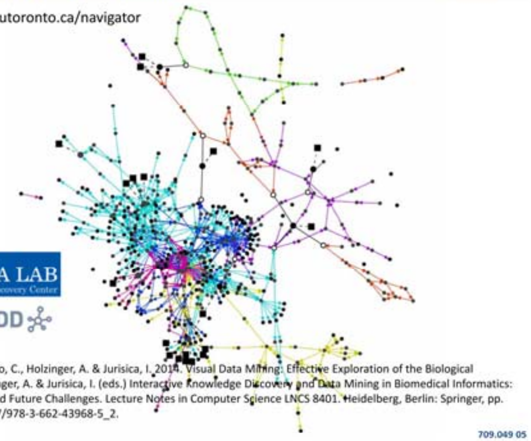
Relationship between
Adrenaline (center black node) and
Dobutamine (top left black node)
Blue: Pharmacological Group
Dark red: Contraindication;
Light red: Condition

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)



Interactive Visual Data Mining

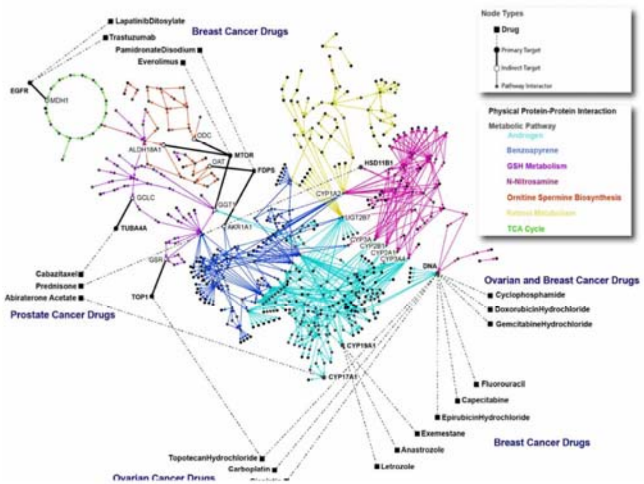
http://ophid.utoronto.ca/navigator



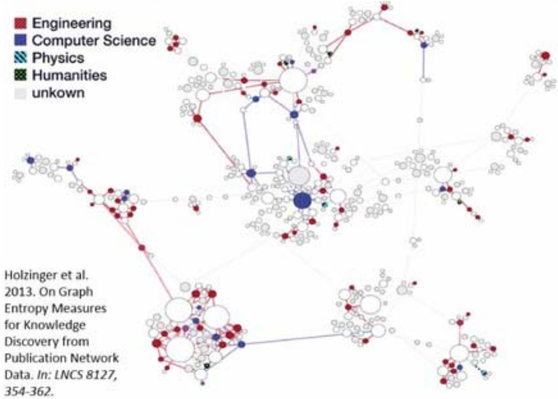
JURISICA LAB
IBM Life Sciences Discovery Center

HCI-KDD

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.



Example: Graph Entropy Measures



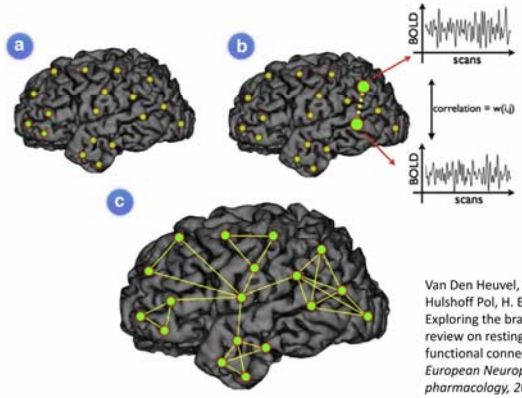
Holzinger et al. 2013. On Graph Entropy Measures for Knowledge Discovery from Publication Network Data. *In: LNCS 8127, 354-362.*

Some selected open problems

- **Problem:** What is the max. number of edges of an Relative Neighborhood Graph in R3 ? 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.

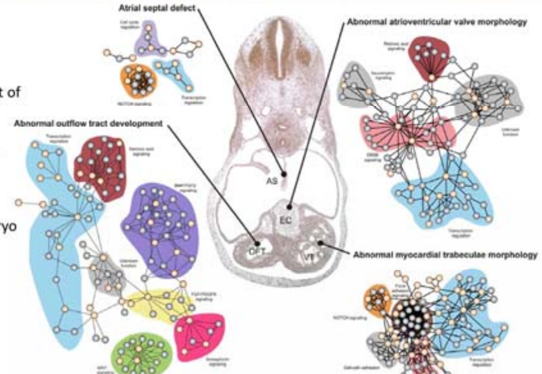
Example: The brain is a complex network



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.*

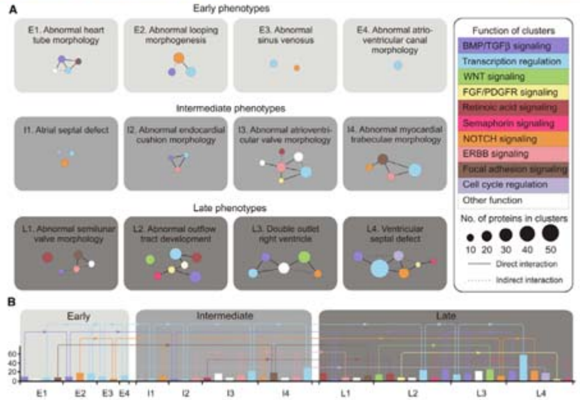
Representative Examples of disease complexes

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



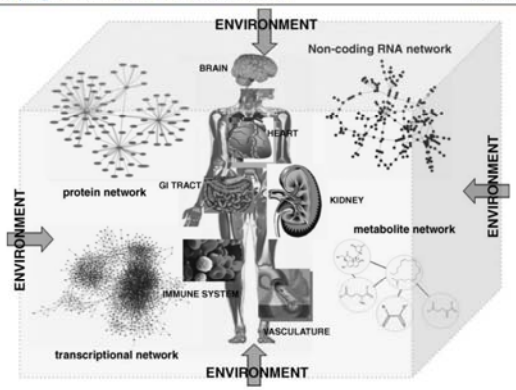
Lage, K. et. al (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular systems biology, 6, 1, 1-9.*

Example: Cell-based therapy



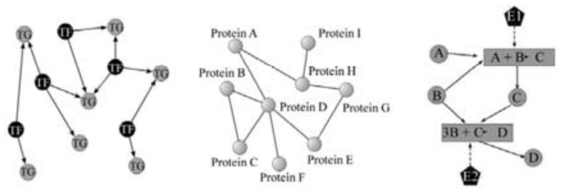
Lage et. al (2010)

Identifying Networks in Disease Research



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.*

Three main types of biomedical networks



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

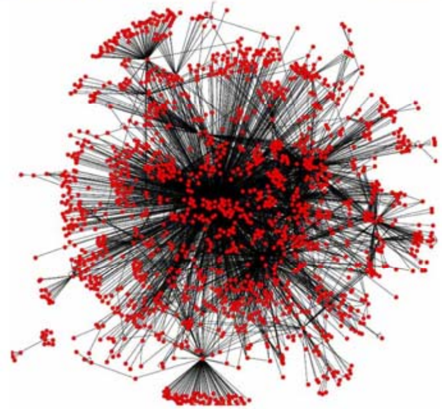
Protein-Protein interaction network

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

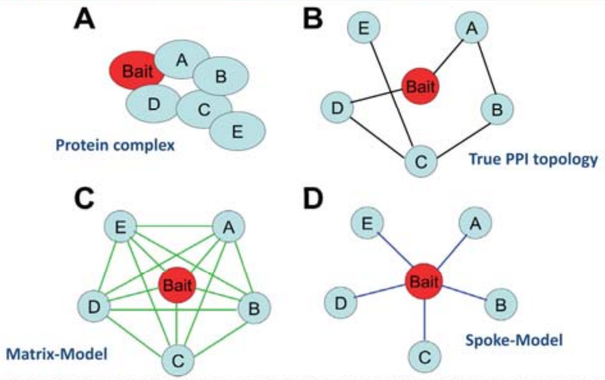
Costa, L. F., Rodrigues, F. A. & Cristino, A. S. (2008) Complex networks: the key to systems biology. *Genetics and Molecular Biology, 31, 3, 591–601.*

Example Transcriptional Regulatory Network

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.*

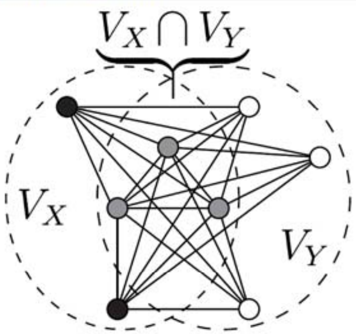


Network Representations of Protein Complexes



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.*

Correlated Motif Mining (CMM)



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.*

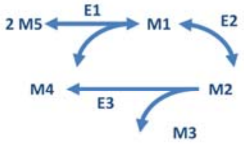
Steepest Ascent Algorithm applied to CMM

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*} ← randomMotifPair()
2: maxsup ← f({X*, Y*}, G)
3: sup ← −∞
4: while maxsup > sup do
5:     {X, Y} ← {X*, Y*}
6:     sup ← maxsup
7:     for all {X', Y'} ∈ N({X, Y}) do
8:         if f({X', Y'}, G) > maxsup then
9:             {X*, Y*} ← {X', Y'}
10:            maxsup ← f({X', Y'}, G)
```

Boyen et al. (2011)

Metabolic Network



	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

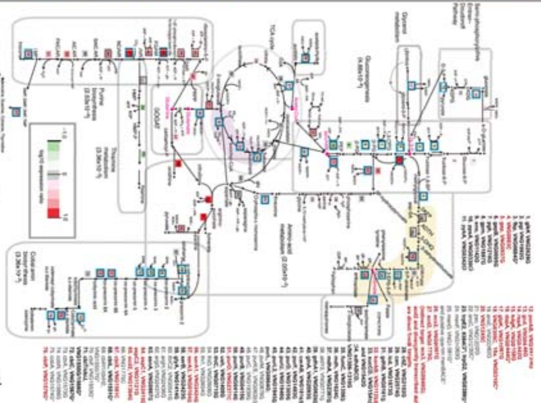
M1	M2
M1	M4
M1	M5
M2	M1
M2	M3
M2	M4
M4	M1
M5	M1

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

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

Metabolic networks are usually big

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.*

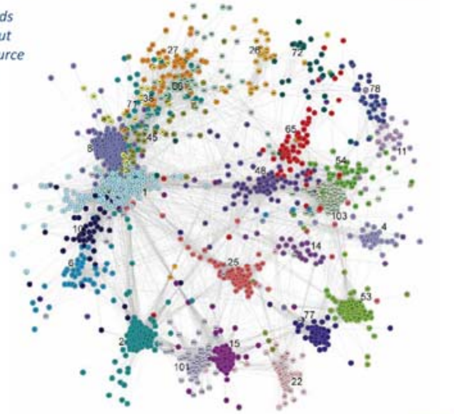


http://www.nature.com/msb/journal/v5/n1/fig_tab/msb200940_F6.html

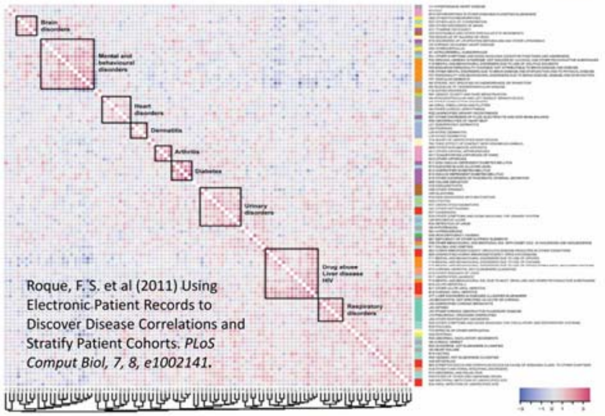
Using EPRs to Discover Disease Correlations

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., Søeby, 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.*



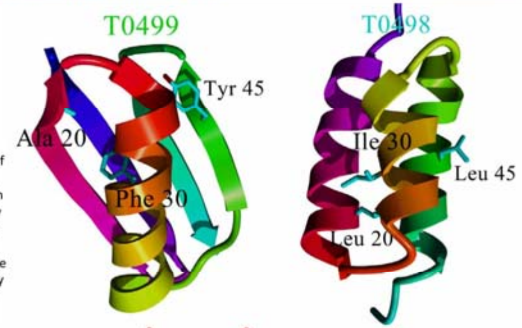
Heatmap of disease-disease correlations (ICD)



Roque, F. S. et al (2011) Using Electronic Patient Records to Discover Disease Correlations and Stratify Patient Cohorts. *PLoS Comput Biol, 7, 8, e1002141.*

Example: ὁμολογέω (homologeo)

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.



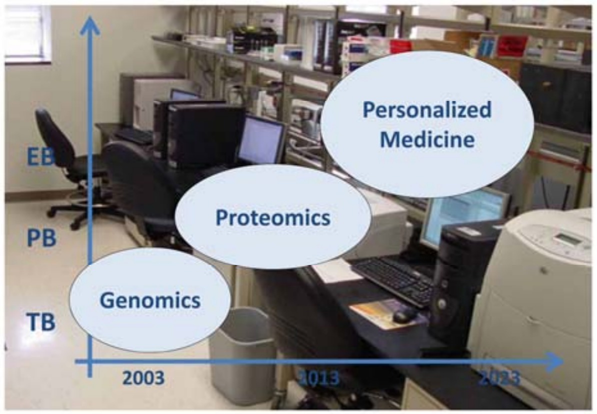
T0499 TTYKLILNLKQAKEEAIKEAVDAGTAEKYFKLIANAKTVEGVWTYKDEIKTFTVTE

T0498 TTYKLILNLKQAKEEAIKELVDAGTAEKYIKLIANAKTVEGVWTLKDEIKTFTVTE

Conclusion

- 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.**

Future Outlook



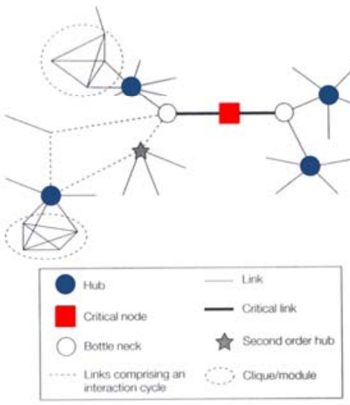
05 Graphs: Basic concepts, metrics and measures

Complex Biological Systems key concepts

- 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

Network Basics on the Example of Bioinformatics

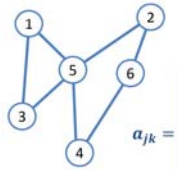
$G(V,E)$ Graph
V ... vertex
E ... edge $\{a,b\}$
$a,b \in V; a \neq b$



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

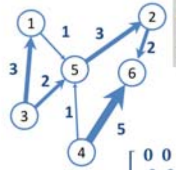
Baby Stuff: Computational Graph Representation

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



$$a_{jk} = \begin{bmatrix} 0&0&1&0&1&0 \\ 0&0&0&0&1&1 \\ 1&0&0&0&1&0 \\ 0&0&0&0&1&1 \\ 1&1&1&1&0&0 \\ 0&1&0&1&0&0 \end{bmatrix}$$

Simple graph, symmetric, binary

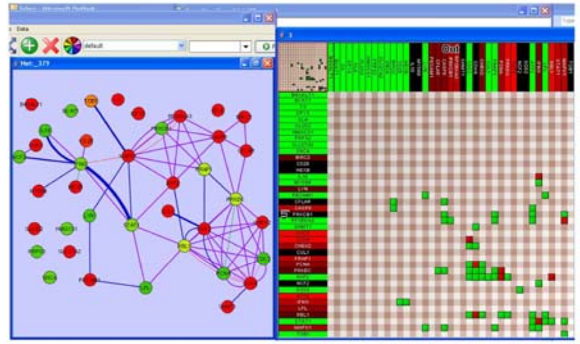


$$w_{jk} = \begin{bmatrix} 0&0&-3&0&1&0 \\ 0&0&0&0&0&3 \\ 3&0&0&0&2&0 \\ 0&0&0&0&1&-5 \\ 1&-2&0&0&-1&0 \\ 0&0&0&0&5&0 \end{bmatrix}$$

Directed and weighted

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

Example: Tool for Node-Link Visualization



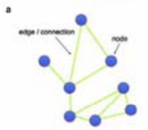
Jean-Daniel Fekete http://wiki.cytoscape.org/InfoVis_Toolkit

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

Some Network Metrics (1/2)

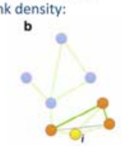
Order = total number of nodes n; **Size** = total number of links (a):

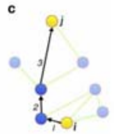
$$\sum_i \sum_j a_{ij}$$



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:

$$C_i = \frac{2t_i}{k(k_i-1)} \qquad C = \frac{1}{n}\sum_i C_i$$





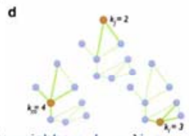
Path length (c) = is the arithmetical mean of all the distances:

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

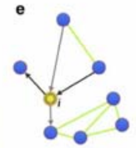
Costa, L. F., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. (2007) Characterization of complex networks: A survey of measurements. *Advances in Physics, 56, 1, 167-242.*

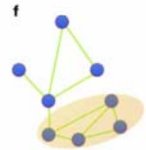
Some Network Metrics (2/2)

- **Centrality** (d) = the level of "betweenness- centrality" of a node I



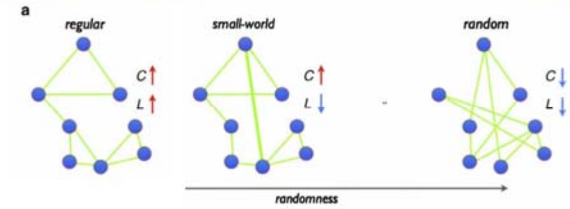
- **Nodal degree** (e) = number of links connecting i to its neighbors: $k_i = \sum_i a_{ij}$

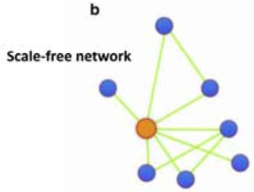




Modularity (f) = describes the possible formation of communities in the network, indicating how strong groups of nodes form relative isolated sub-networks within the full network (refer also to Slide 5-8).

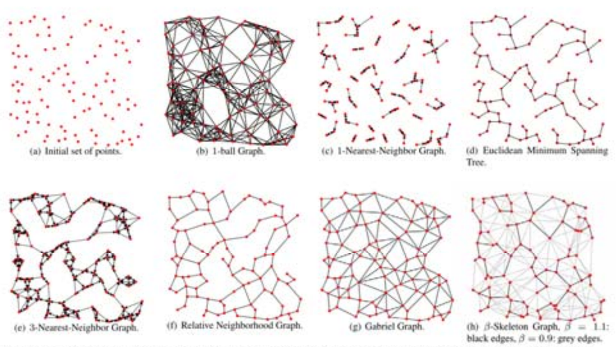
Network Topologies





Van Heuvel & Hulshoff (2010)

Slide 5-15 Graphs from Point Cloud Data Sets

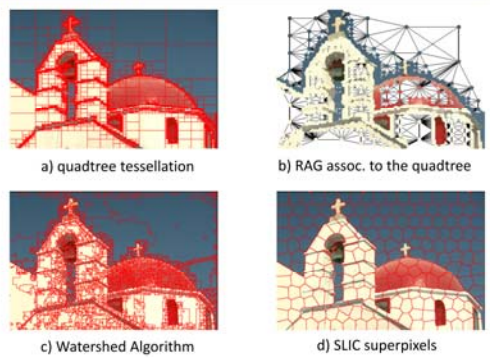


Lézoray, O. & Grady, L. 2012. Graph theory concepts and definitions used in image processing and analysis. *In: Lézoray, O. & Grady, L. (eds.) Image Processing and Analysing With Graphs: Theory and Practice. Boca Raton (FL): CRC Press, pp. 1-24.*

Finally a practical example

06 Example: How do you get point cloud data from natural images?

Graphs from Images



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.*

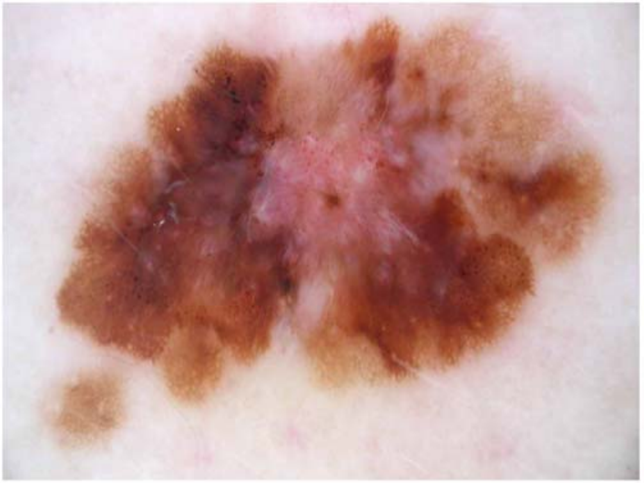
Example Watershed Algorithm

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

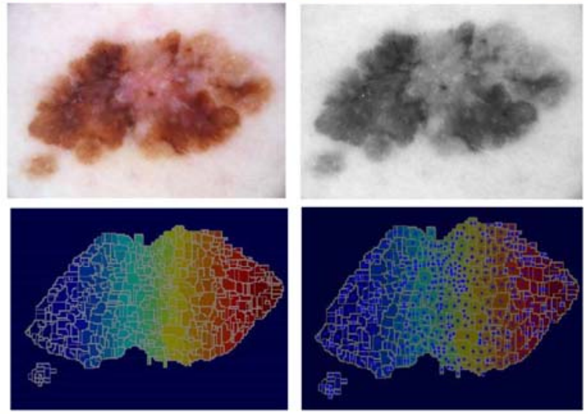
```
procedure ShortestPathWatershed;
INPUT: lower complete digital grey scale image G = (V, E, im) with cost function cost.
OUTPUT: labelled image lab on V.
#define WSHED 0                      (* label of the watershed pixels *)
(* Uses distance image dist. On output, dist[v] = im[v], for all v ∈ V. *)

for all v ∈ V do        (* Initialize *)
   lab[v] ← 0 ; dist[v] ← ∞
end for
for all local minima m_i do
   for all v ∈ m_i do
      lab[v] ← i ; dist[v] ← im[v]        (* initialize distance with values of minima *)
   end for
end for
while V ≠ ∅ do
   u ← GetMinDist(V)        (* find u ∈ V with smallest distance value dist[u] *)
   V ← V\{u}
   for all v ∈ V with (u, v) ∈ E do
      if dist[u] + cost[u, v] < dist[v] then
         dist[v] ← dist[u] + cost(u, v)
         lab[v] ← lab[u]
      else if lab[v] ≠ WSHED and dist[u] + cost[u, v] = dist[v] and lab[v] ≠ lab[u] then
         lab[v] = WSHED
      end if
   end for
end while
```

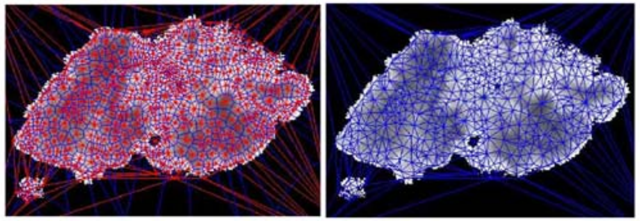
Meijster, A. & Roerdink, J. B. A proposal for the implementation of a parallel watershed algorithm. Computer Analysis of Images and Patterns, 1995. Springer, 790-795.



Graphs from Images: Watershed + Centroid



Slide 5-20 Graphs from Images: Voronoi <> Delaunay



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.

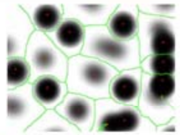
Are graphs better than feature vectors ?

- 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

Watershed methods



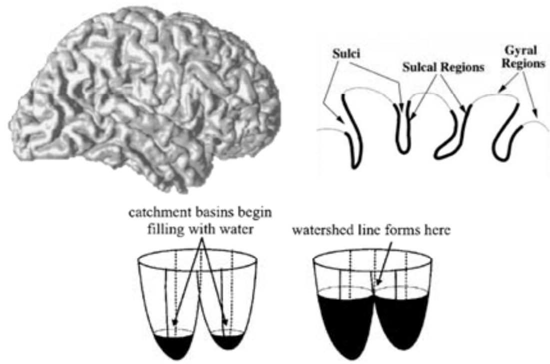
- Topographic maps => landscapes with height structures
- Segmentation into regions of pixels
- Assuming drops of water raining on the map
- Following paths of descent
- Lakes called catchment basins
- Also possible: Flooding based
- Needs Topographical distance measures (MST)

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.

Watershed 4 Steps

- 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

Example: Watershed Segmentation of the human cortex



http://iacl.ece.jhu.edu/~prince/ws/

Finally a practical example

07 Graphical Model Learning

Learning Graphical Models from data

- Remember: GM are a marriage between probability theory and graph theory and provide a tool for dealing with our two grand challenges in the biomedical domain:

Uncertainty and complexity

- The learning task is two-fold:
  1) Learning unknown probabilities
  2) Learning unknown structures

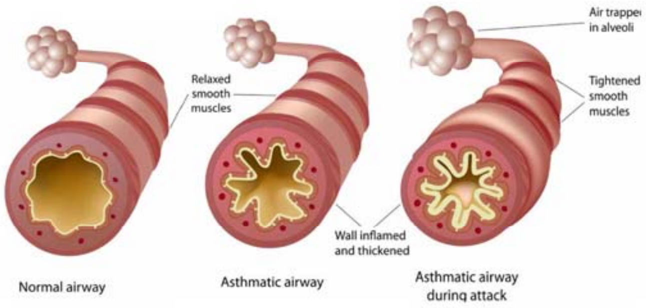
Jordan, M. I. 1998. Learning in graphical models, Springer

Learning the Structure of GM from data

1) Test if a distribution is decomposable with regard to a given graph.
   - This is the most direct approach. It is not bound to a graphical representation,
   - It can be carried out w.r.t. other representations of the set of subspaces to be used to compute the (candidate) decomposition of a given distribution.
2) Find a suitable graph by measuring the strength of dependences.
   - This is a heuristic, but often highly successful approach, which is based on the frequently valid assumption that in a conditional independence graph an attribute is more strongly dependent on adjacent attributes than on attributes that are not directly connected to them.
3) Find an independence map by conditional independence tests.
   - This approach exploits the theorems that connect conditional independence graphs and graphs that represent decompositions.
   - It has the advantage that a single conditional independence test, if it fails, can exclude several candidate graphs. Beware, because wrong test results can thus have severe consequences.

Borgelt, C., Steinbrecher, M. & Kruse, R. R. 2009. Graphical models: representations for learning, reasoning and data mining, John Wiley & Sons.

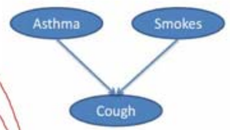
Who of you smokes?



Beasley, R. 1998. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The Lancet, 351, (9111), 1225-1232, doi:http://dx.doi.org/10.1016/S0140-6736(97)07302-9.

Example for Graphical Model Learning

Patient	J46	Tussis	Smoker
Florian	1	1	0
Tamas	0	0	0
Matthias	1	0	0
Benjamin	0	1	1
Dimitrios	0	1	0
...			
...			



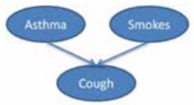
Florian	0	?	?

Rows are independent during learning and inference!

Florian	0	0.3	0.2

Relational Representation Learning and Prediction

- Asthma can be hereditary
- Friends may have similar smoking habits
- Augmenting graphical model with relations between the entities – Markov Logic



2.1 Asthma ⇒ Cough

3.5 Smokes ⇒ Cough

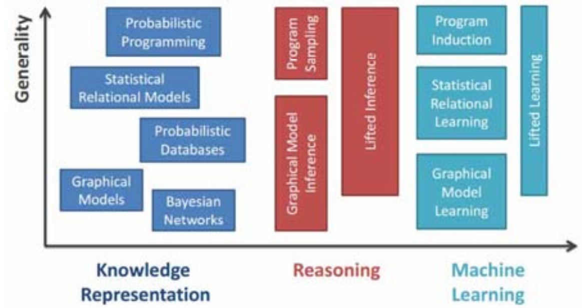
2.1 Asthma(x) ⇒ Cough(x)

3.5 Smokes(x) ⇒ Cough(x)

1.9 Smokes(x) ∧ Friends(x,y) ⇒ Smokes(y)

1.5 Asthma (x) ∧ Family(x,y) ⇒ Asthma (y)

Knowledge Representation > Reasoning > Learning



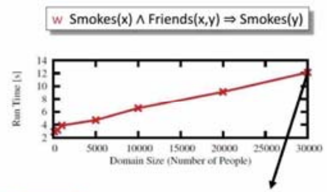
Example for probabilistic rule learning, in which probabilistic rules are learned from probabilistic examples: The ProbFOIL+ Algorithm solves this problem by combining the principles of the rule learner FOIL with the probabilistic Prolog called ProbLog, see: De Raedt, L., Dries, A., Thon, I., Van Den Broeck, G. & Verbeke, M. 2015. Inducing probabilistic relational rules from probabilistic examples. International Joint Conference on Artificial Intelligence (IJCAI).

Conclusion and Future Challenges

Future Outlook

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**

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.



~~Big data~~

Big models

Learns a model over 900,030,000 random variables

Thank you!

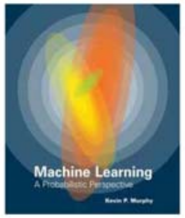
Questions

Sample Questions

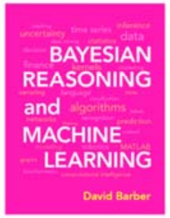
- What is the primary idea of a graphical model learning algorithm?
- Where do graphs come from in the medical domain?
- Where do decision trees originally come from?
- What are probabilistic graphical models?
- Why is the topic "reasoning under uncertainty" so important for the health domain?
- Why was MYCIN not a success in the clincial domain?
- What was the core essence in MYCIN?
- What is the principle of GAMUTS?
- Which two types of decisions do clinicians execute?
- What is the goal of network medicine?
- What is a true PPI topology?
- Why are structural homologies so important?
- What is the vision of personalized medicine?
- What does robustness in the context of complex biological systems mean?
- How do you get point cloud data from a natural image?
- Why is graphical model learning so interesting for medical problems?

Appendix

Recommended Books

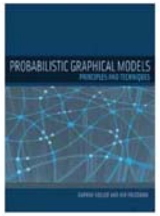


Murphy, K. P. 2012. Machine learning: a probabilistic perspective, MIT press.



Barber, D. 2012. *Bayesian reasoning and machine learning*, Cambridge University Press.

http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/181115.pdf



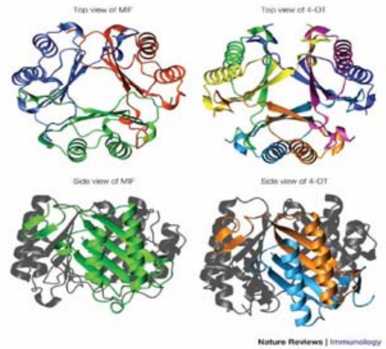
Koller, D. & Friedman, N. 2009. Probabilistic graphical models: principles and techniques, MIT press.

Key Problems

- Automated Machine Learning algorithms need much training data – focus is on adjusting model parameters without fully **understanding the data** that the learning algorithm is modeling [1]
- Curse of dimensionality [2] – need for privacy and **anonymization** [3] (see lecture 11)
- **Weakly structured data** [4]

[1] Smith, M. R., Martinez, T. & Giraud-Carrier, C. 2014. An instance level analysis of data complexity. *Machine learning,* 95, (2), 225-256.
[2] Friedman, J. H. 1997. On bias, variance, 0/1—loss, and the curse-of-dimensionality. Data mining and knowledge discovery, 1, (1), 55-77.
[3] Aggarwal, C. C. On k-anonymity and the curse of dimensionality. Proceedings of the 31st international conference on Very large data bases VLDB, 2005. 901-909
[4] Holzinger, A., Stocker, C. & Dehmer, M. 2014. Big Complex Biomedical Data: Towards a Taxonomy of Data. In: CCIS 455. Berlin Heidelberg: Springer pp. 3-18.

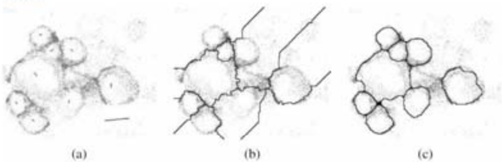
Example from Immunology: Structural Homology



Calandra, T. & Roger, T. 2003. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol, 3,* ***791-800.***

Slide 5-19: Watershed Principle

- Catchment basins:
  - treating an image as a height field or landscape, regions where the rain would flow into the same lake



- Start flooding from local minima, and label ridges wherever differently evolving components meet

U.S. Patent May 7, 2002 Sheet 5 of 11 US 6,384,826 B1

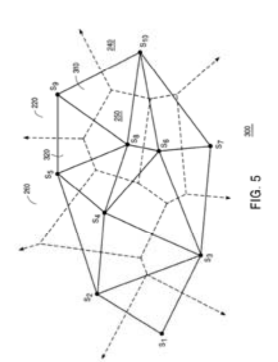


FIG. 5

U.S. Patent May 7, 2002 Sheet 6 of 11 US 6,384,826 B1

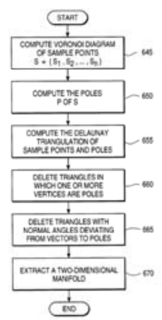
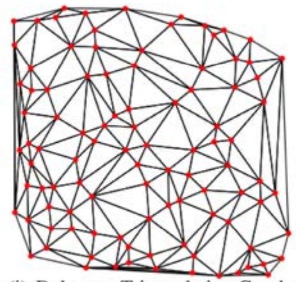


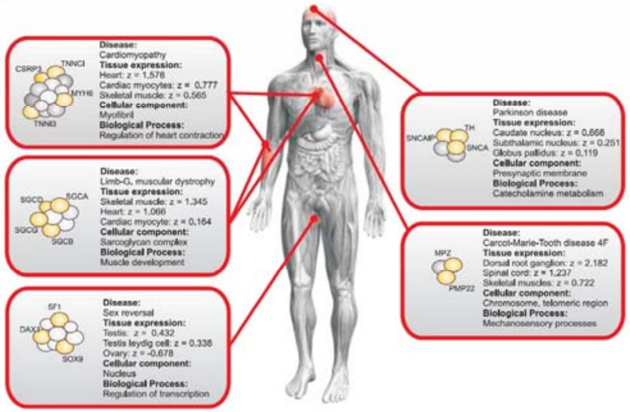
FIG. 6

Slide 5-15 Graphs from Images: Voronoi <> Delauney

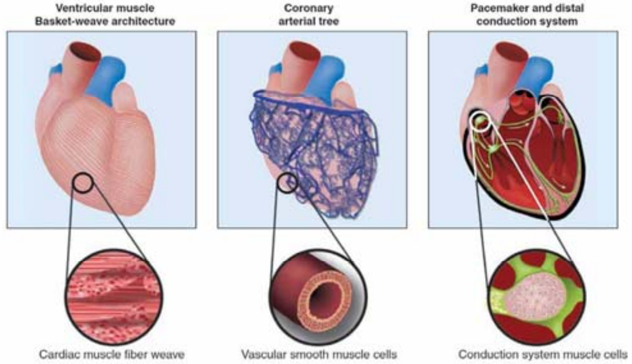
(a) Initial set of points.



(i) Delaunay Triangulation Graph.

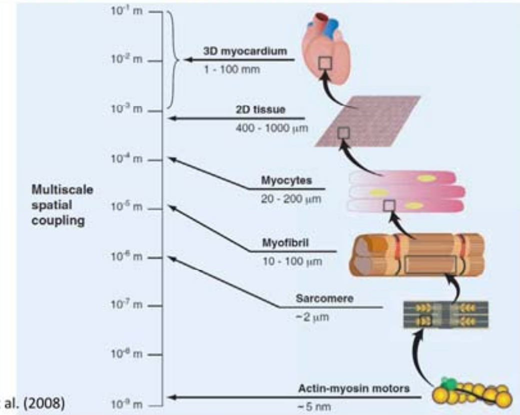


Example: Cell based therapy (1) (Heart transplantation)



Chien, K. R., Domian, I. J. & Parker, K. K. (2008) Cardiogenesis and the complex biology of regenerative cardiovascular medicine. *Science, 322, 5907, 1494.*

Example: Cell based therapy (2) (Heart transplantation)

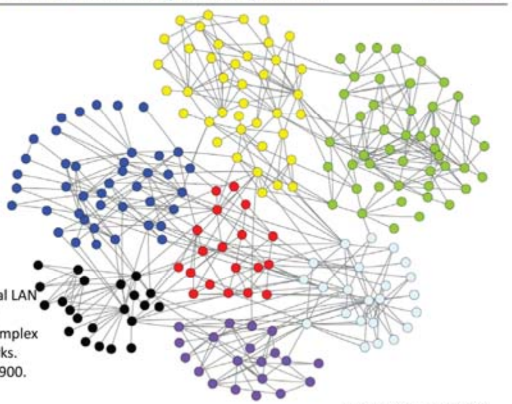


Chien et al. (2008)

Example: Network Generated by Gene Duplication

High Modularity (Modularity = 0.6717, Scaled Modularity = 29); Different colors represent different modules identified by Guimera and Amaral's algorithm [28].

Guimera R, Amaral LAN (2005) Functional cartography of complex metabolic networks. Nature 433: 895–900.



Wang & Zhang (2007)

Genome-Phenome association in complex diseases

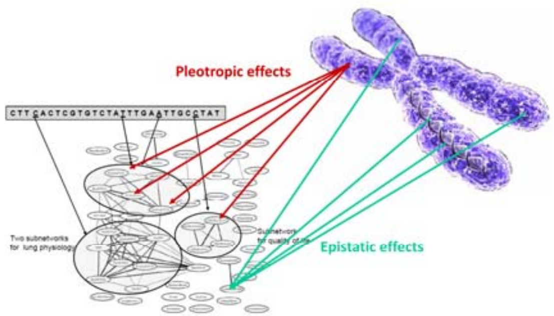
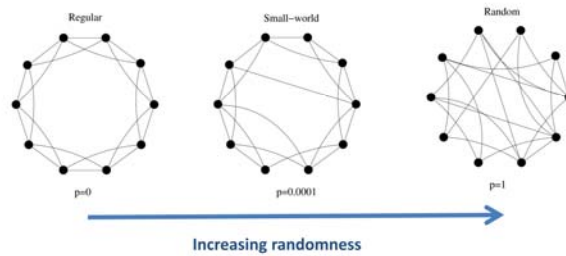


Image credit to Eric Xing, Carnegie Mellon University, Pittsburgh

Small-World Networks

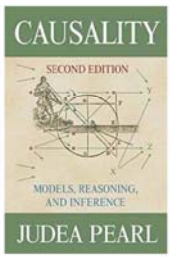


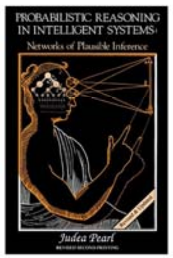
29.000 citations ...

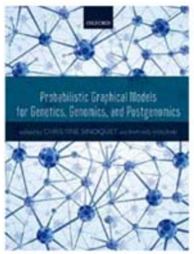
Watts, D. J. & Strogatz, S. (1998) Collective dynamics of small-world networks. *Nature, 393, 6684, 440-442.*

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

Additional Reading







Pearl, J. 2009. Causality: Models, Reasoning, and Inference (2nd Edition), Cambridge, Cambridge University Press.

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

Sinoquet, C. & Mourad, R. 2014. Probabilistic Graphical Models for Genetics, Genomics, and Postgenomics, OUP Oxford.