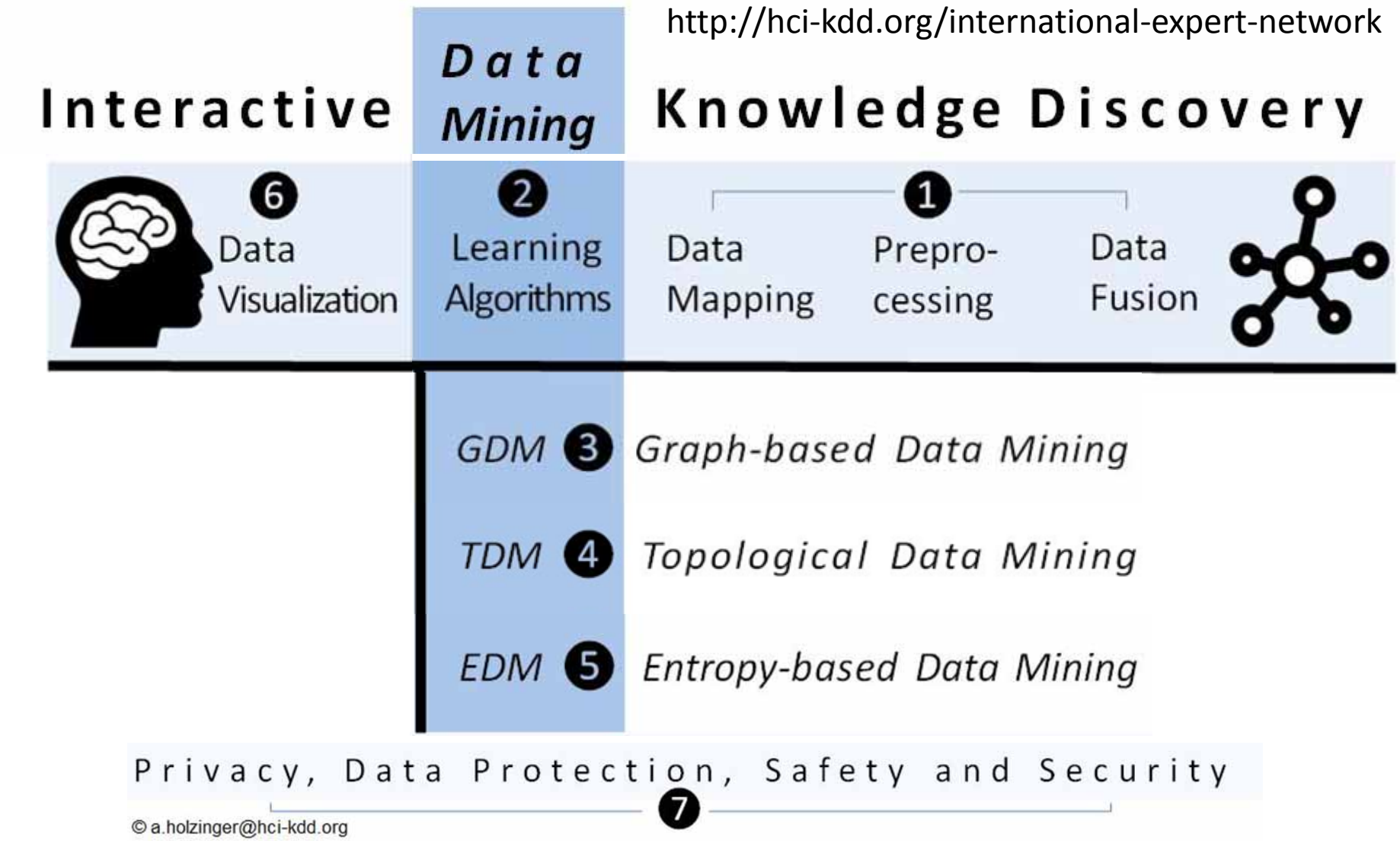
185.A83 Machine Learning for Health Informatics
2016S, VU, 2.0 h, 3.0 ECTS
Week 19 11.05.2016 17:00-20:00

Probabilistic Graphical Models
Part 1: From Knowledge Representation to Graph Model Learning

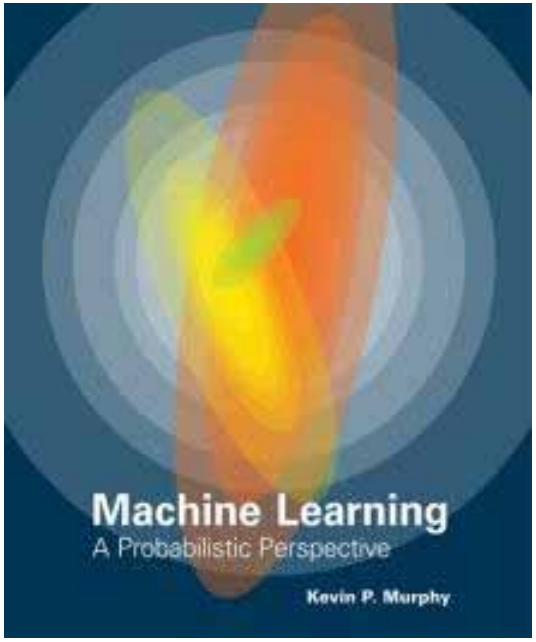
a.holzinger@hci-kdd.org
http://hci-kdd.org/machine-learning-for-health-informatics-course

ML needs a concerted effort fostering integrated research



Please always note that for sustainable solving of biomedical problems we need a concerted effort of different disciplines fostering an integrated research

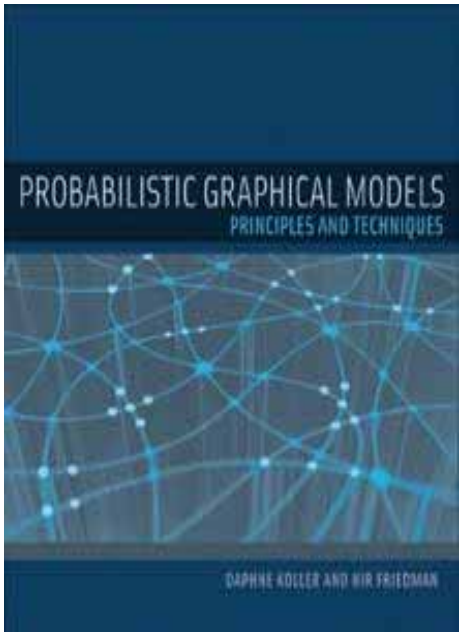
Recommended Books



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

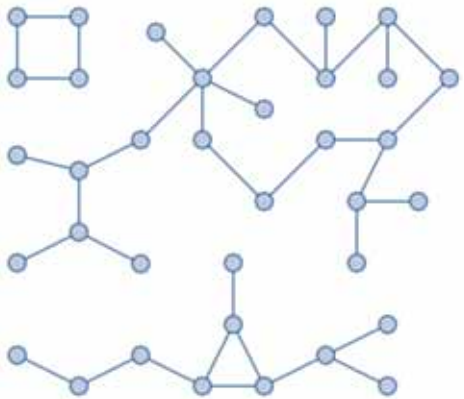


Koller, D. & Friedman, N. 2009. Probabilistic graphical models: principles and techniques, MIT press.

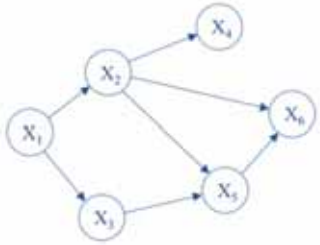
Red thread through this lecture

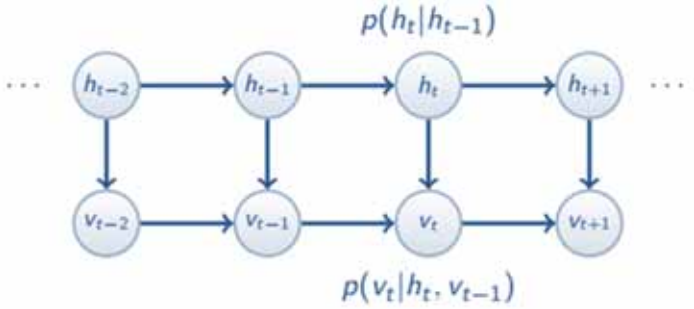
- 1) Reasoning under uncertainty
- 2) Where do graphs come from?
- 3) What are the challenges?
- 4) Knowledge Representation in Network Medicine
- 5) Review of basic metrics and measures
- 6) Practical Example: How do we get PCD from natural images?
- 7) Graphical Model Learning

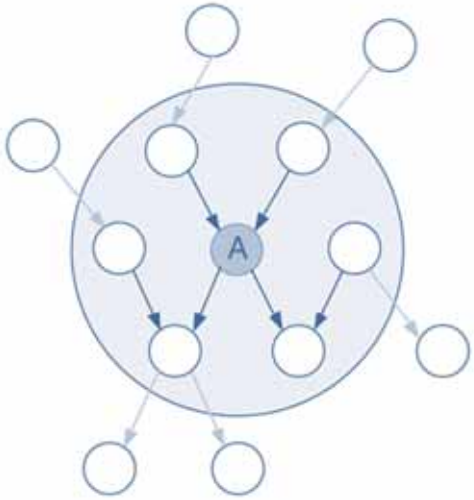
Quiz: Which concepts can you identify …



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





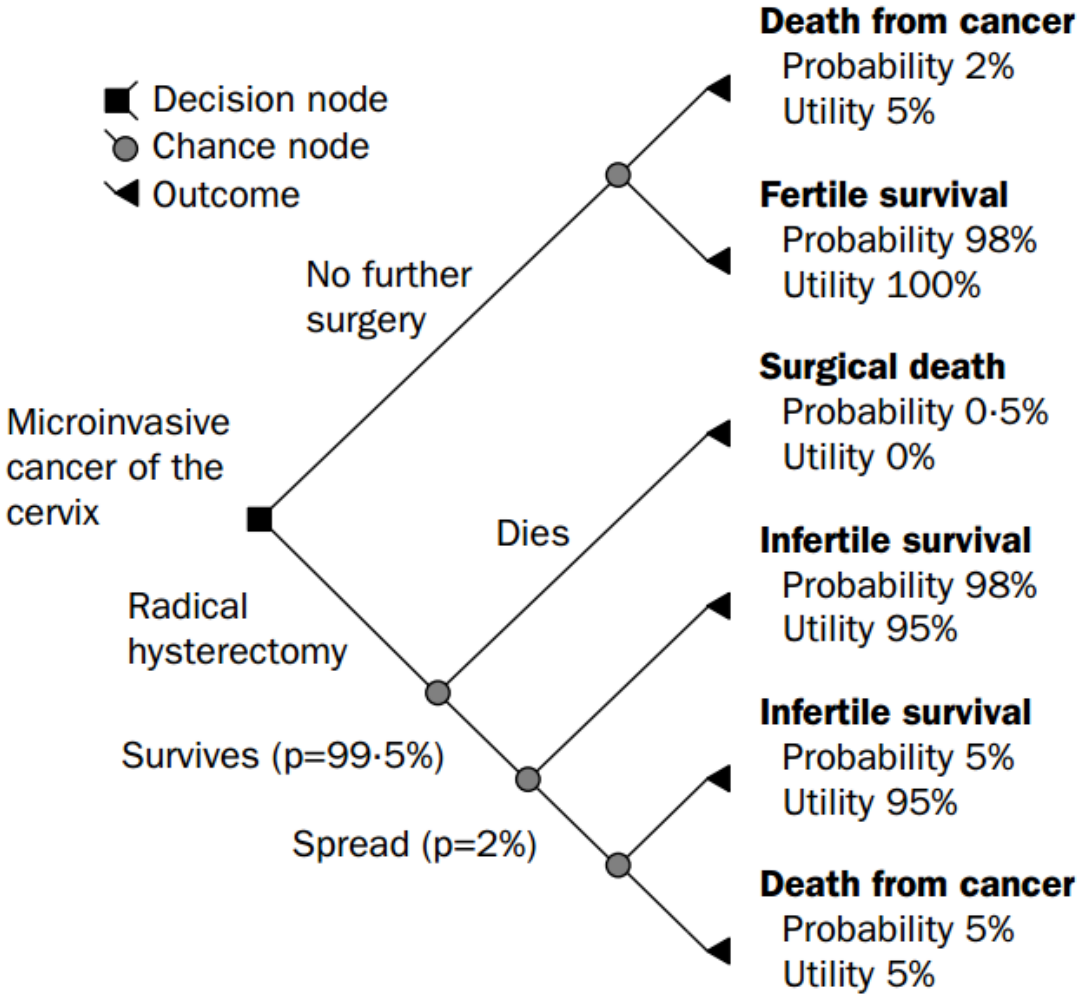


$$P(\mathbf{x}) = \prod_{i \in V} P_i(x_i) \prod_{(i,j) \in E} \frac{P_{i,j}(x_i, x_j)}{P_i(x_i) P_j(x_j)}$$

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

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**

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 $E\ (U|d)$

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

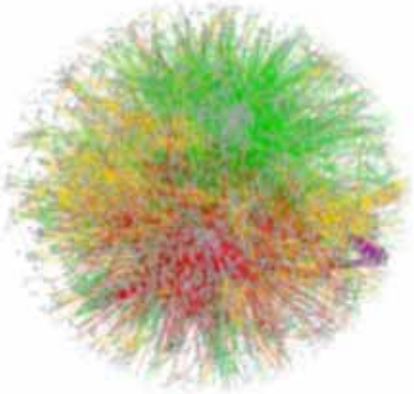
$$E(U \mid d) = \sum_{x_1,\ldots,x_n} P(x_1, \ldots, x_n \mid d) U(x_1, \ldots, x_n, d)$$

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.

- PGM can be seen as a combination between
- **Graph Theory + Probability Theory + Machine Learning**
- One of the most exciting advancements in AI 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



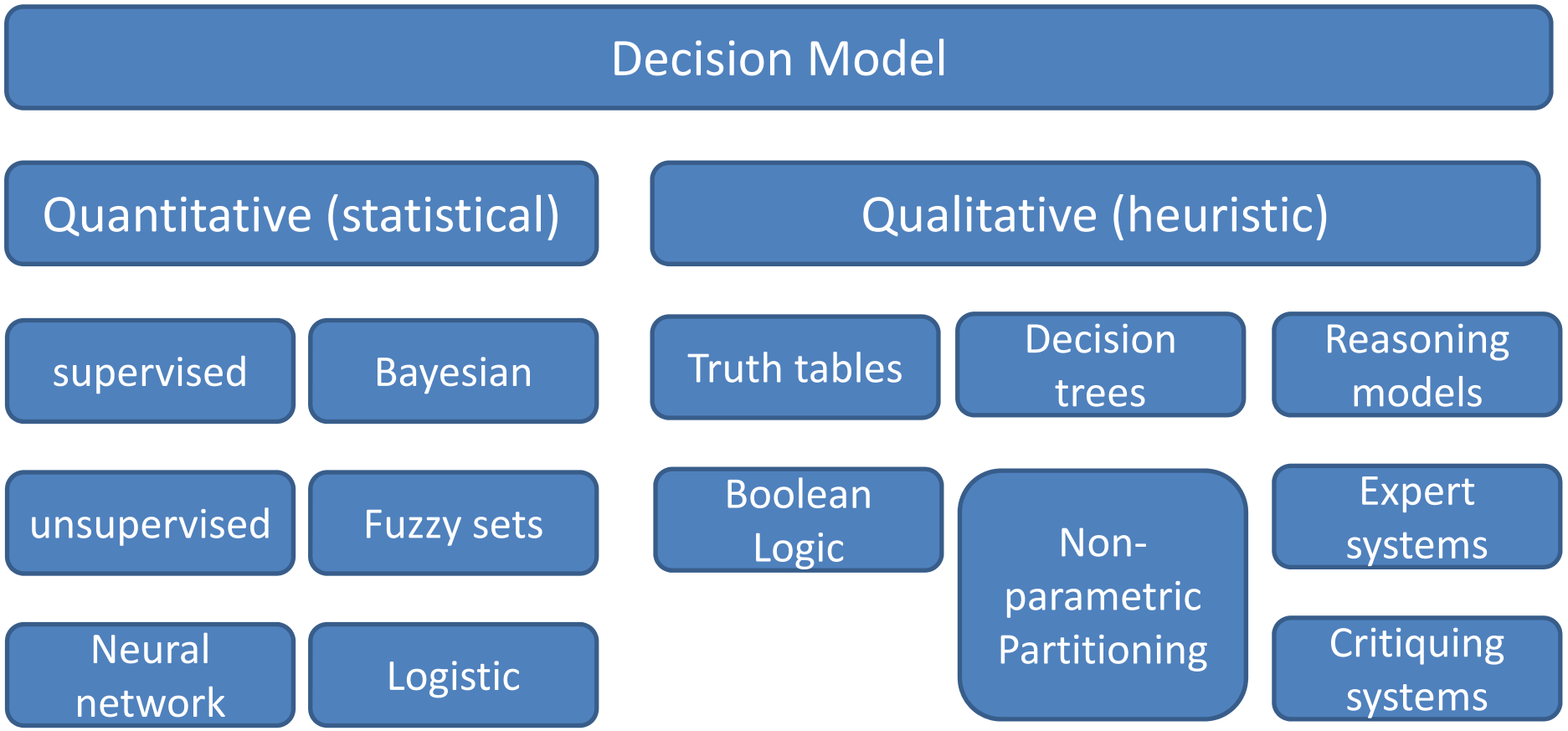
Key Challenges

- 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.

1) Reasoning under Uncertainty

Remember: Taxonomy of Decision Support Models



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 – 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 – rule based system - certainty factors

- 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 -> h is 100 % true
- CF[h] = –1 -> h is 100% false

Original Example from MYCIN

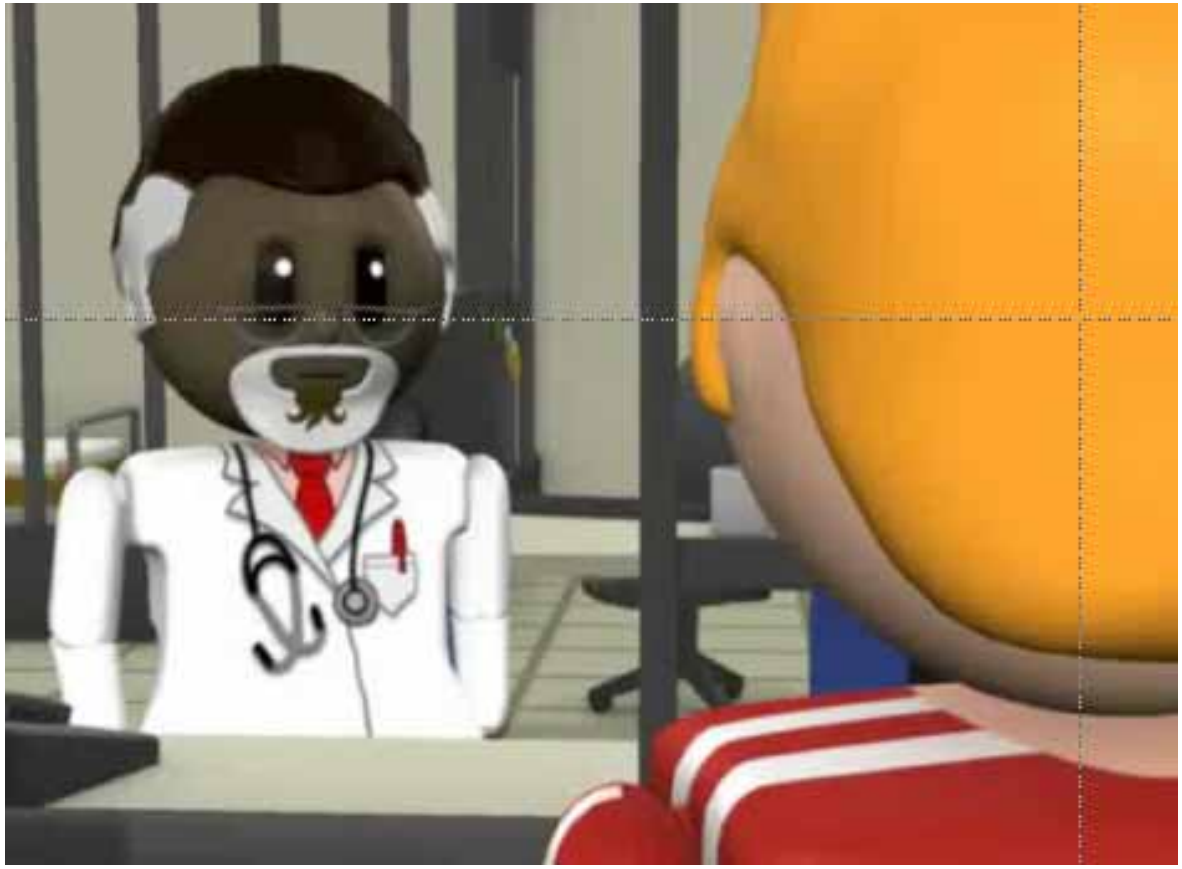
h_1 = The identity of ORGANISM-1 is streptococcus
h_2 = PATIENT-1 is febrile
h_3 = The name of PATIENT-1 is John Jones

$CF[h_1,E] = .8$	:	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

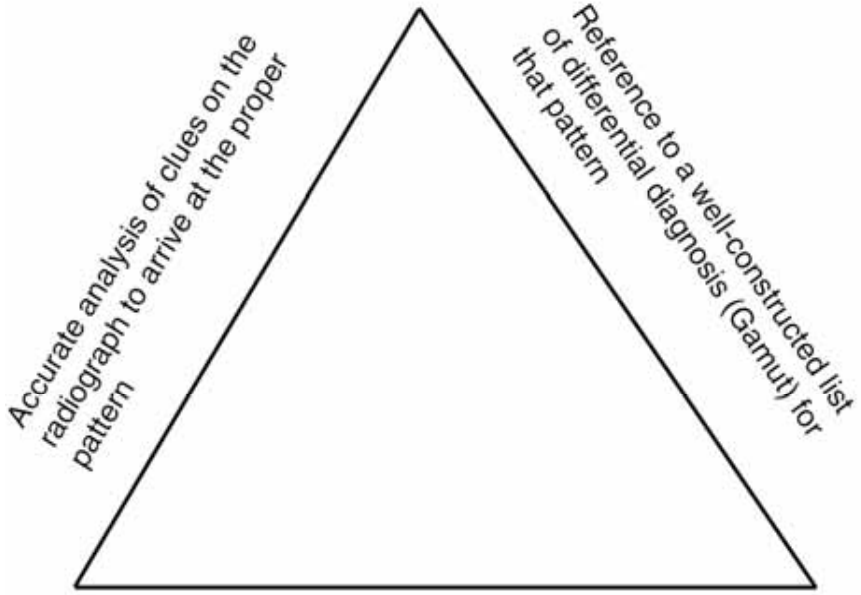
Shortliffe, E. H. & Buchanan, B. G. (1984) *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley.*

MYCIN was *no* success in the clinical practice

https://www.youtube.com/watch?v=IVGWM0CKNWA ("real nurse triage")



Gamuts: Triangulation to find diagnoses



Correlation of radiographic findings and Gamut with patients' clinical and lab findings to arrive at the most likely diagnosis

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

Example - Gamuts in Radiology

REEDER AND FELSON'S

GAMUTS IN RADIOLOGY

GAMUT G-25
EROSIVE 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

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

Remember: 2 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?
Why are they so interesting?

Leonhard Euler 1736 …

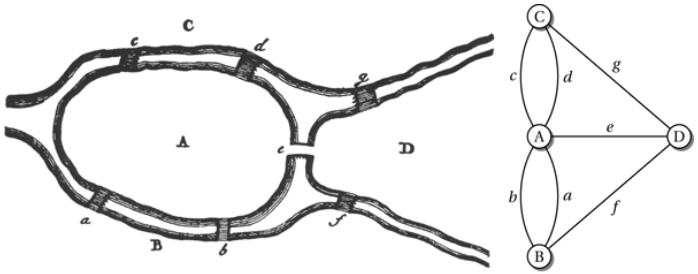
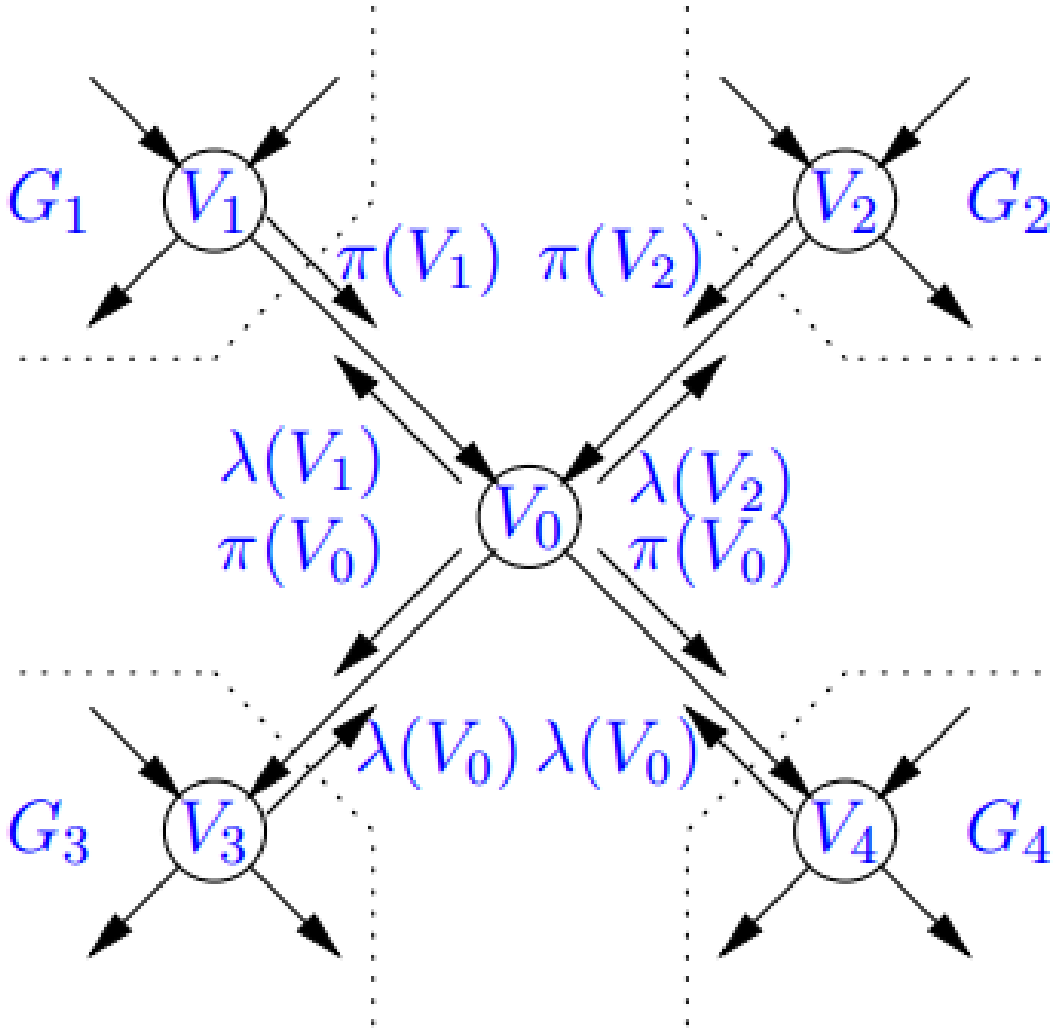


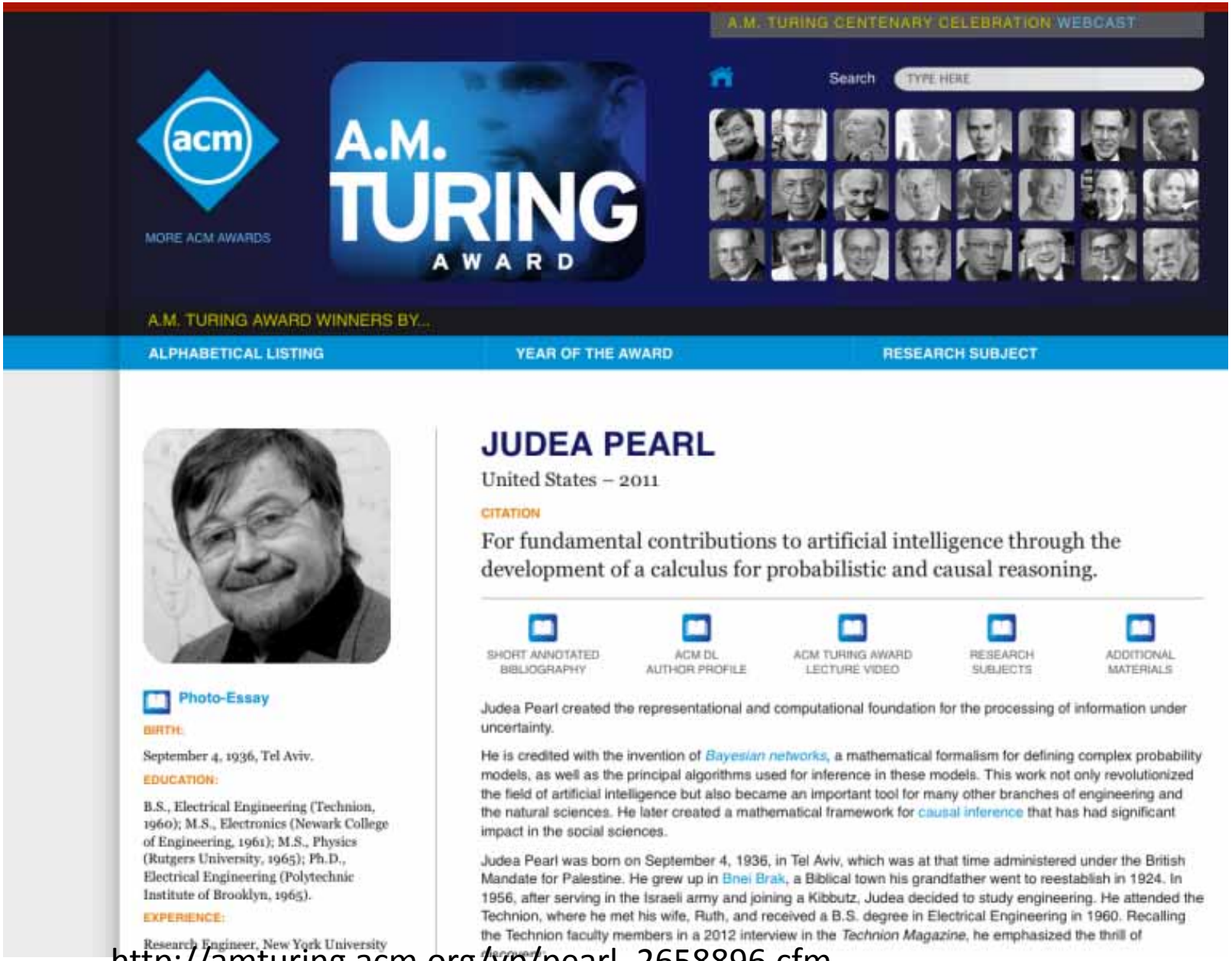
Image from https://people.kth.se/~carlofi/teaching/FEL3250-2013/courseinfo.html

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

Scientific Background on the Nobel Prize in Chemistry 2013

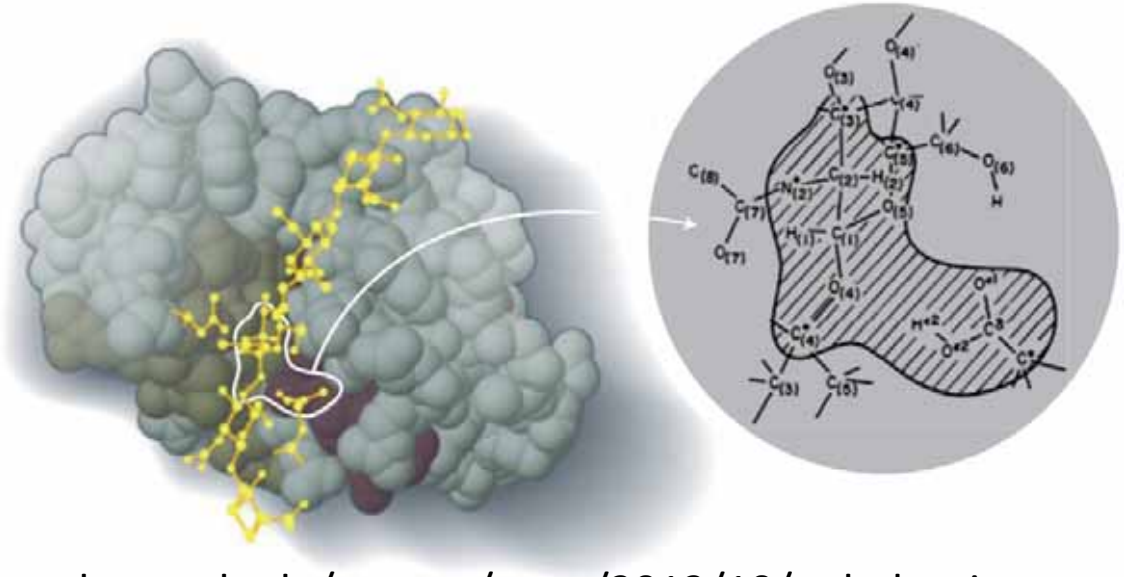
DEVELOPMENT OF MULTISCALE MODELS FOR COMPLEX CHEMICAL SYSTEMS

Photo: A. Mahmoud
Martin Karplus
Prize share: 1/3

Photo: A. Mahmoud
Michael Levitt
Prize share: 1/3

Photo: A. Mahmoud
Arieh Warshel
Prize share: 1/3

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



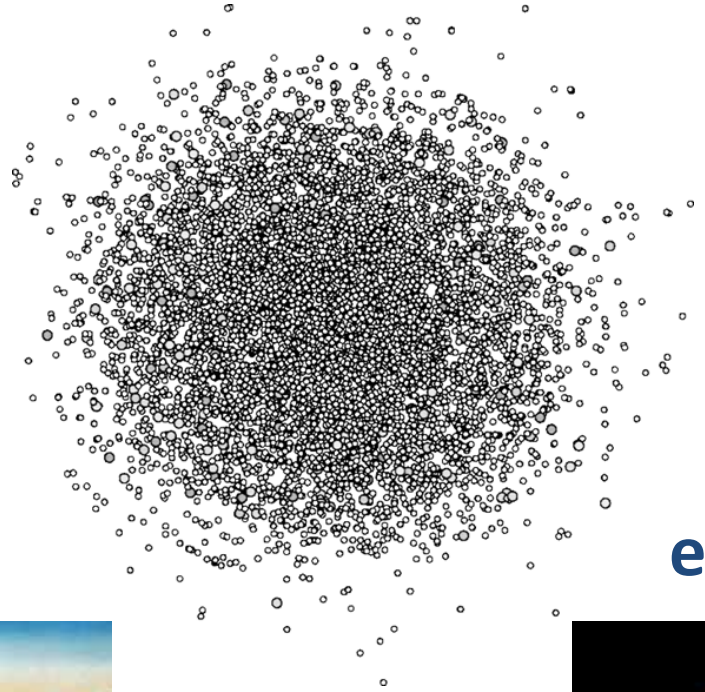
http://news.harvard.edu/gazette/story/2013/10/nobel_prize_awarded_2013/

First Question: Where does graphs come from?

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

- **Graphs as nonparametric basis**
- we learn the structure from samples and infer
- flat vector data, e.g. similarity graphs
- encoding structural properties (e.g. smoothness, independence, …)

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



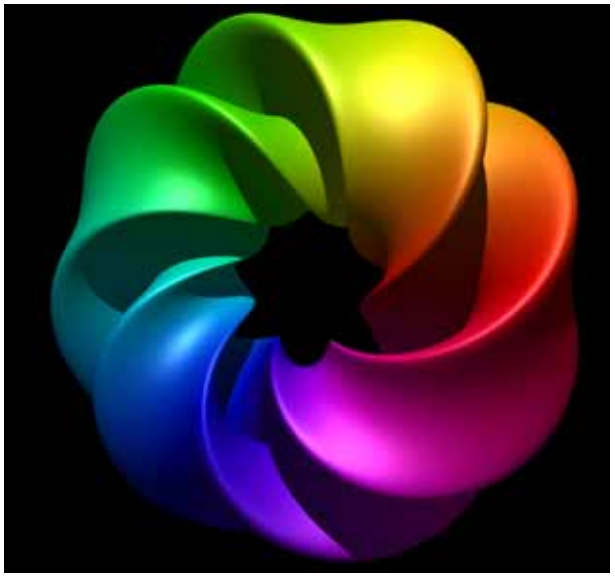
Time

e.g. Entropy

Dali, S. (1931) The persistence of memory

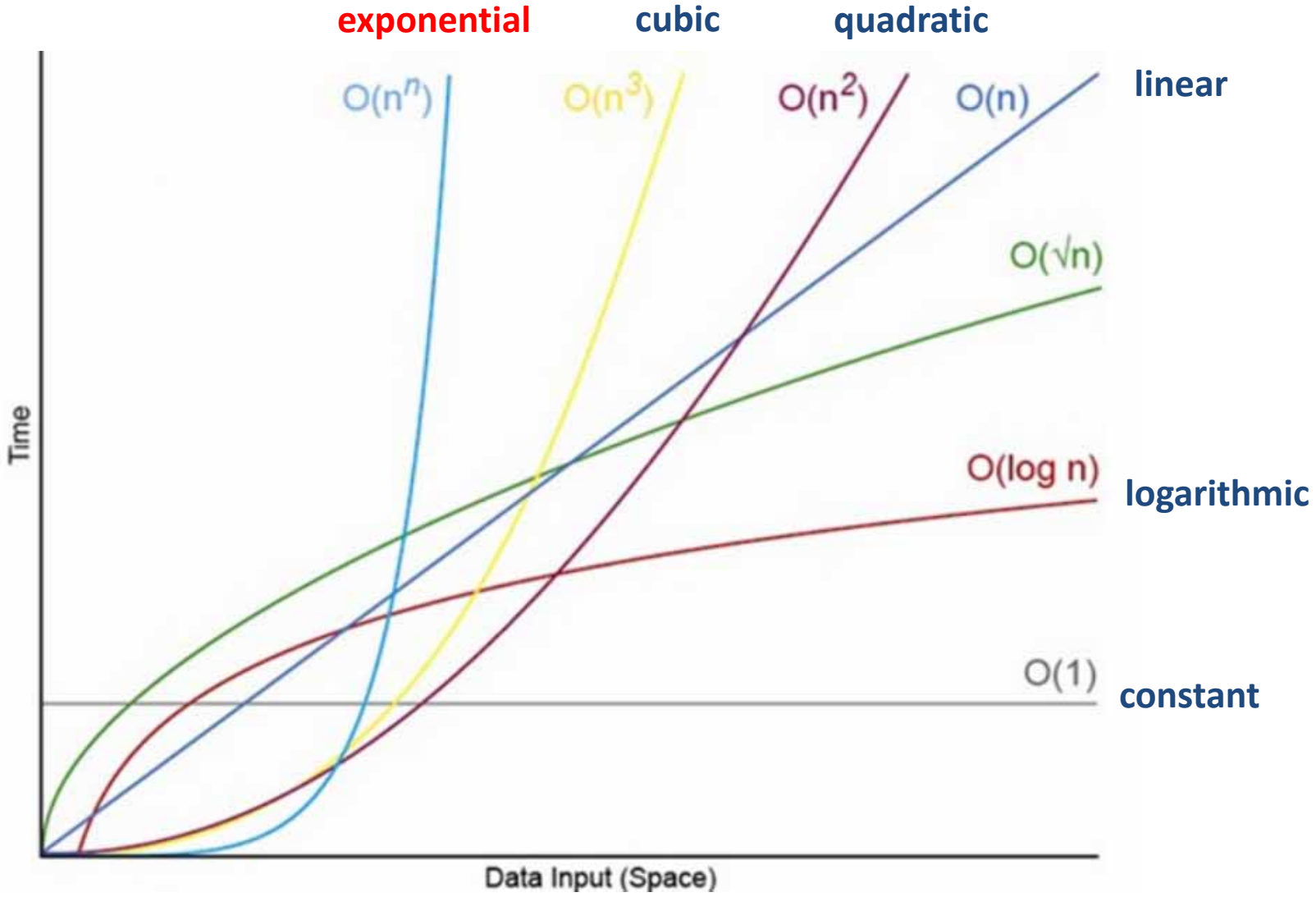
Space

e.g. Topology



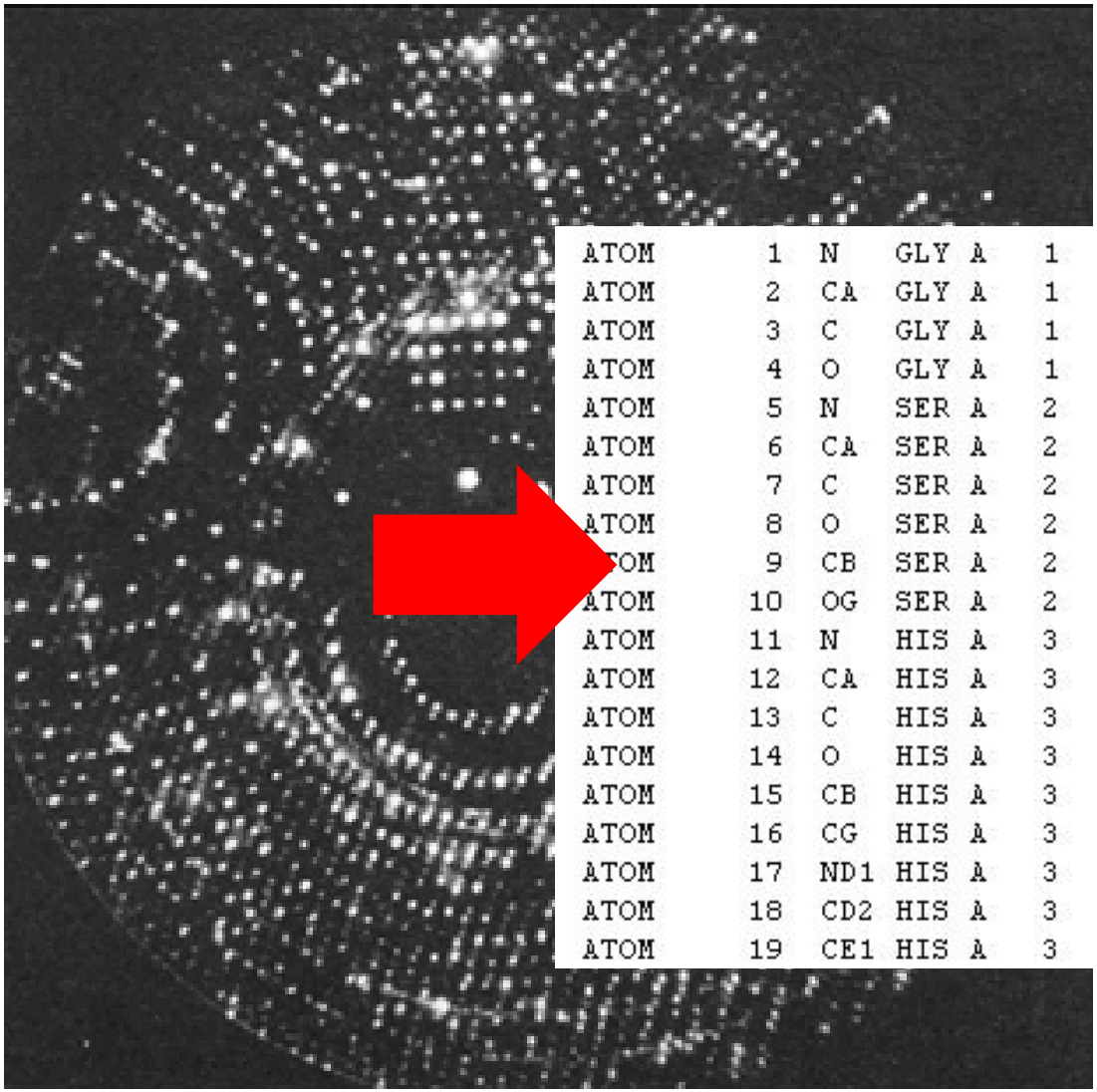
Bagula & Bourke (2012) Klein-Bottle

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

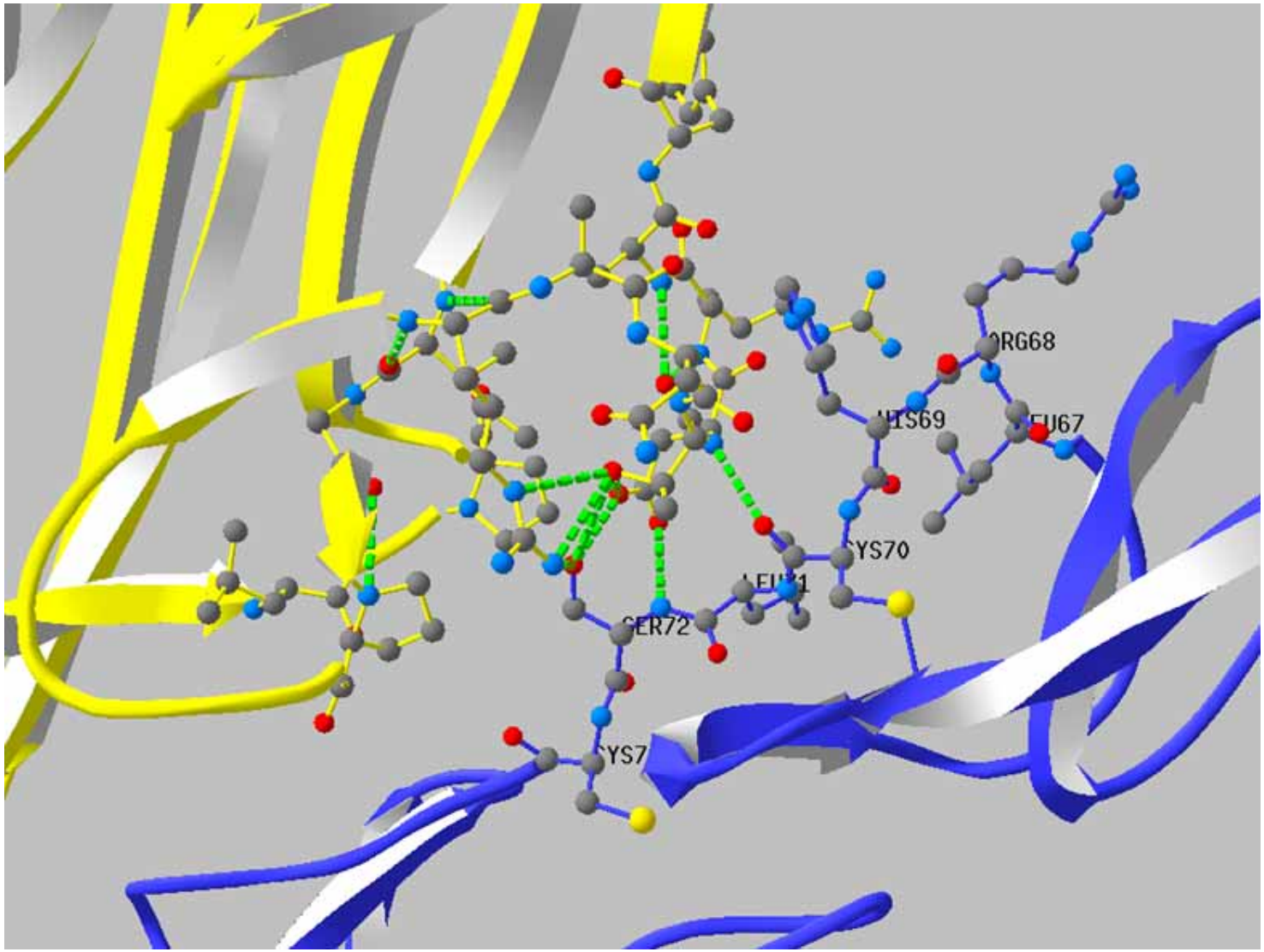
Our World in Data – Microscopic Structures



```
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
```

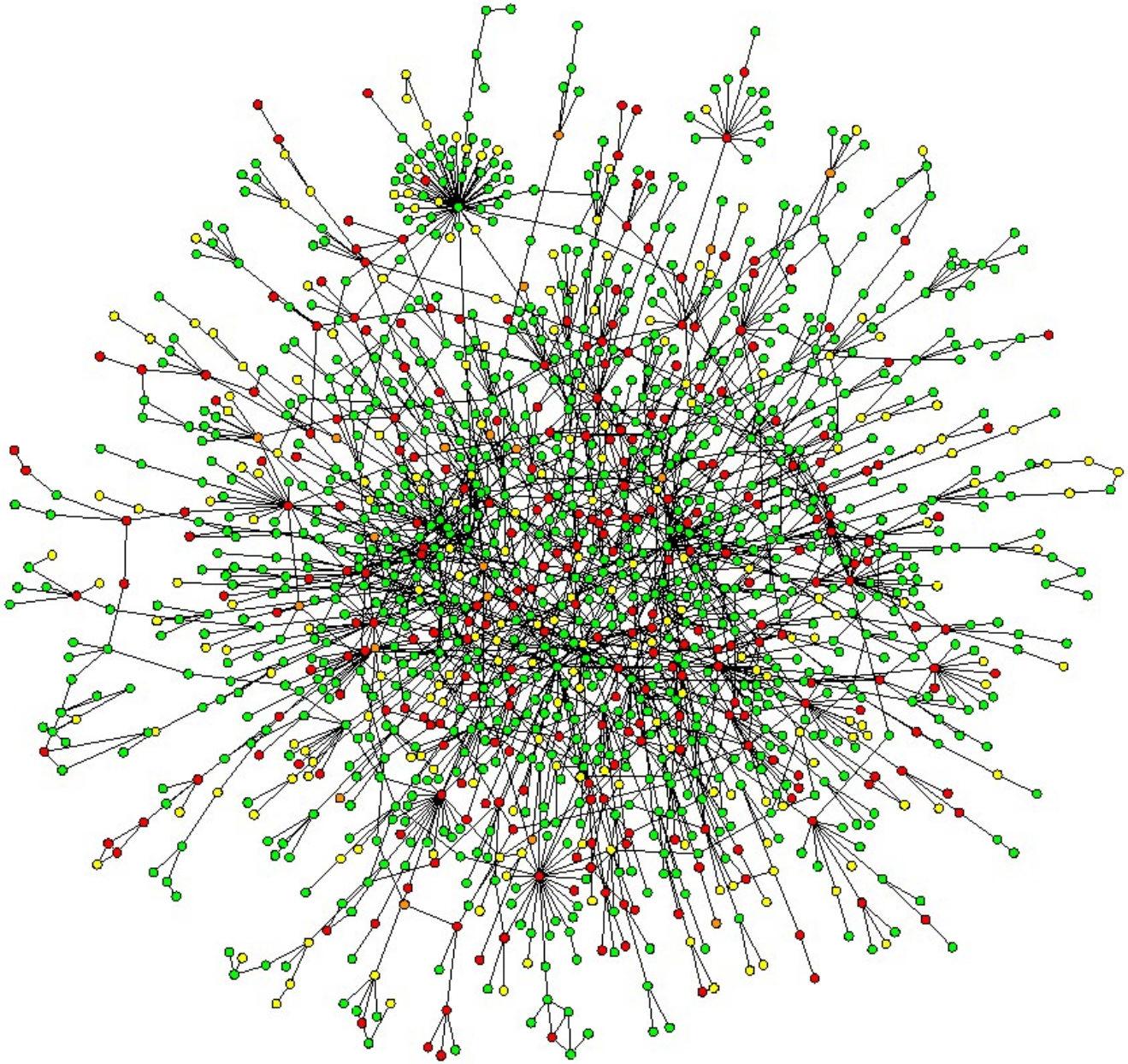
Wiltgen, M. & Holzinger, A. (2005) Visualization in Bioinformatics: Protein Structures with Physicochemical and Biological Annotations. In: *Central European Multimedia and Virtual Reality Conference. Prague, Czech Technical University (CTU), 69-74*

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

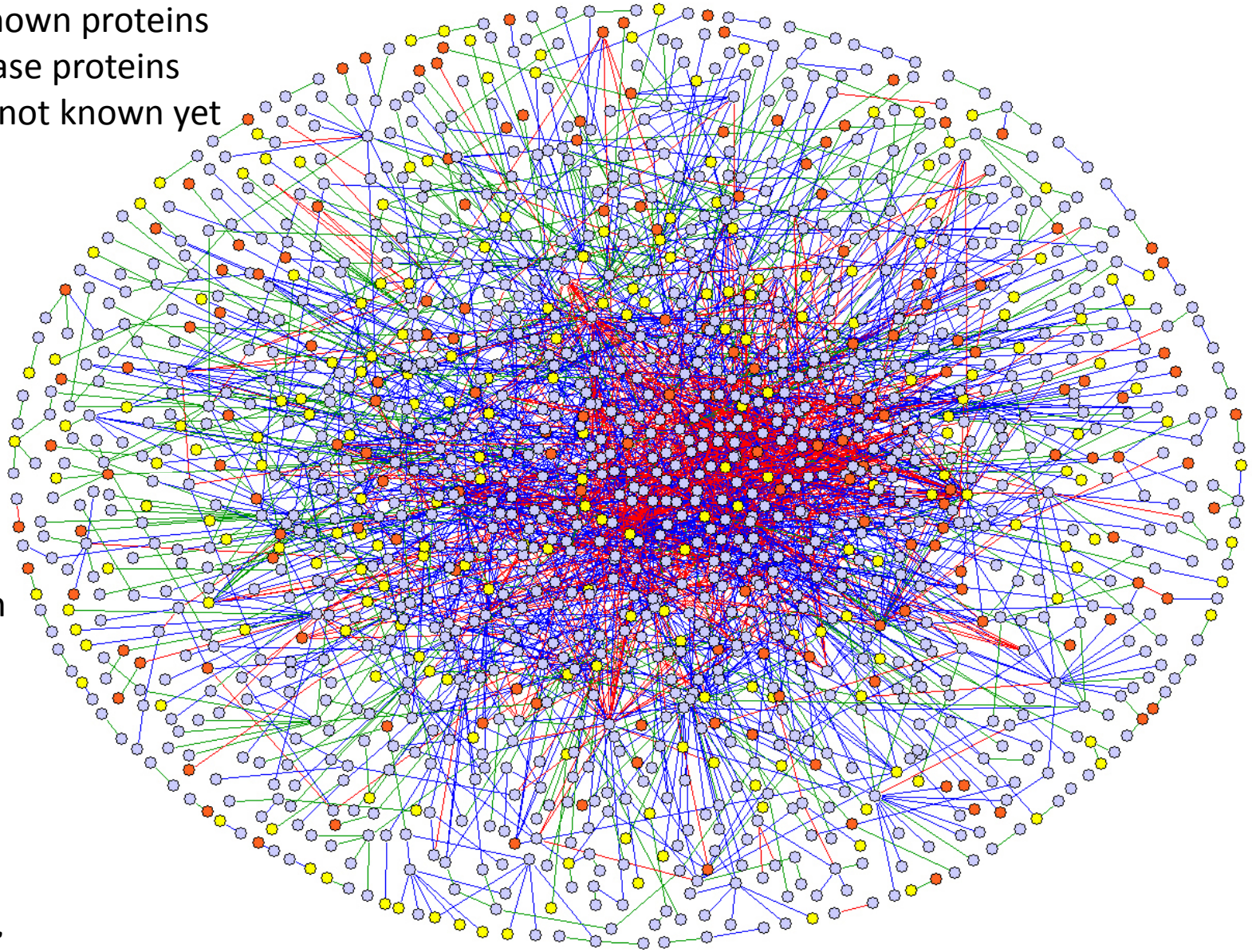


Nodes = proteins
Links = physical interactions (bindings)
Red Nodes = lethal
Green Nodes = non-lethal
Orange = slow growth
Yellow = not known

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

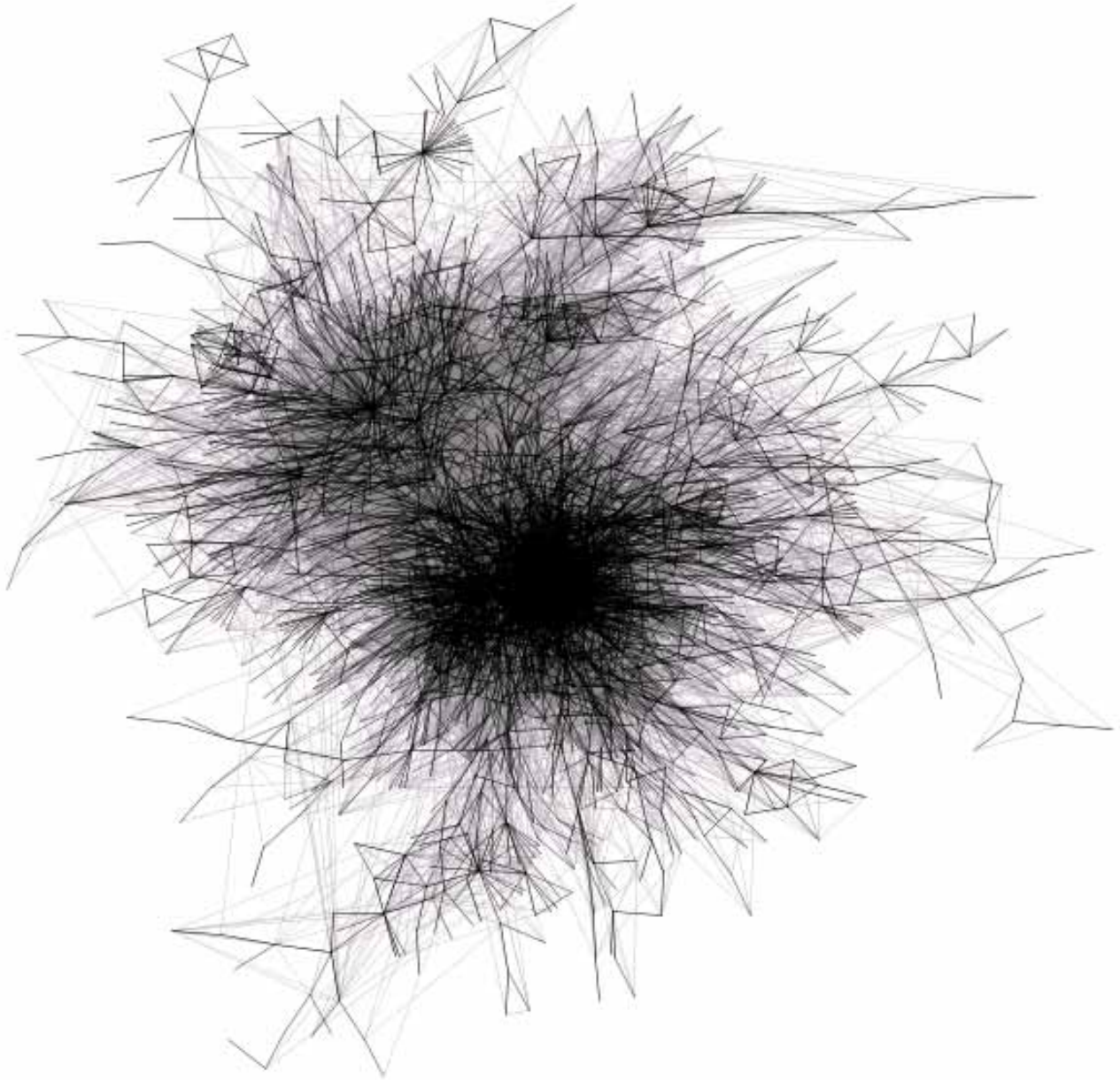
First human protein-protein interaction network

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



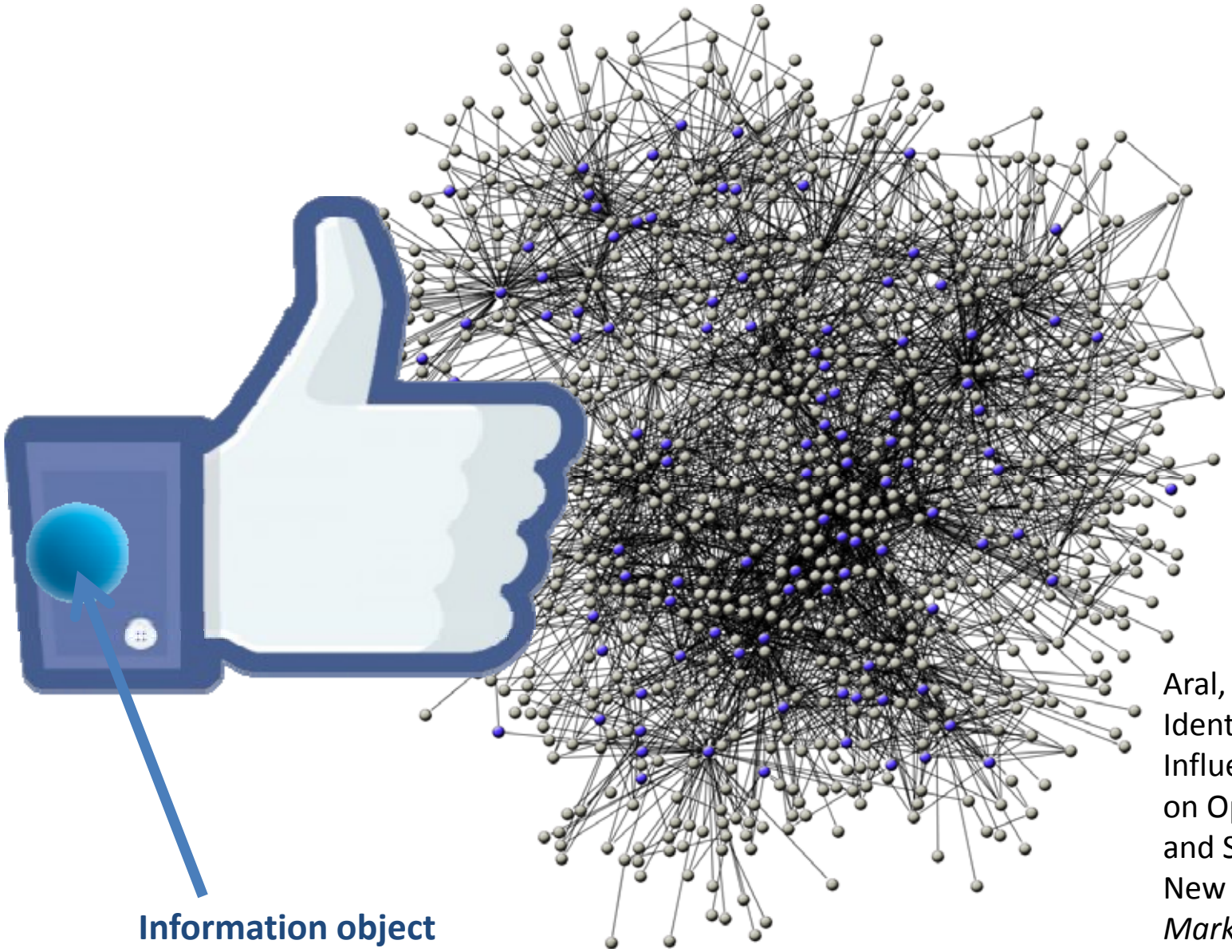
Stelzl, U. et al. (2005) A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. *Cell, 122, 6, 957-968*.

Non-Natural Network Example: Blogosphere



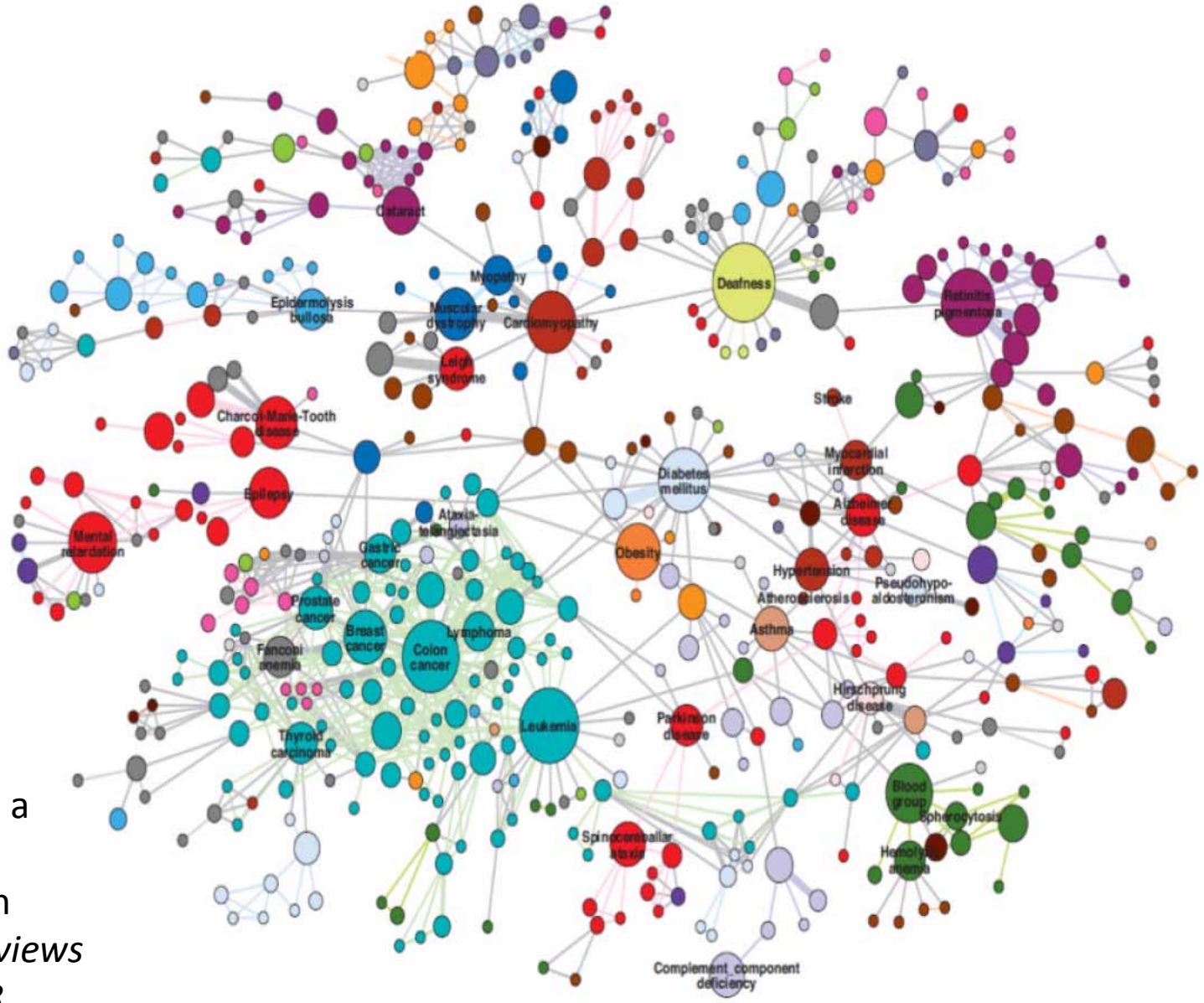
Hurst, M. (2007), Data Mining: Text Mining, Visualization and Social Media. Online available: http://datamining.typepad.com/data_mining/2007/01/the_blogosphere.html, last access: 2011-09-24

Social Behavior Contagion Network



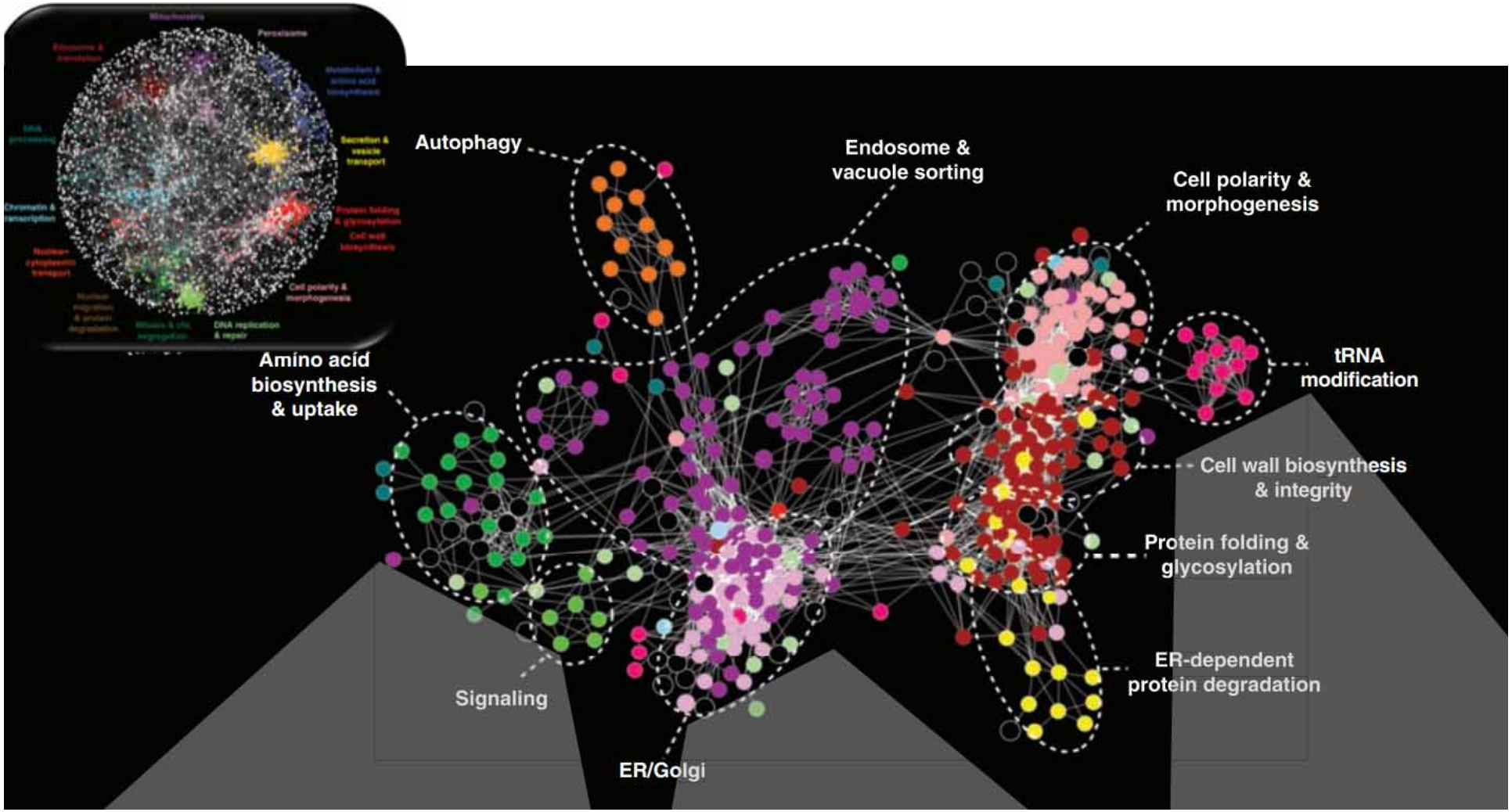
Aral, S. (2011) Identifying Social Influence: A Comment on Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science, 30, 2, 217-223.*

Human Disease Network -> Network Medicine



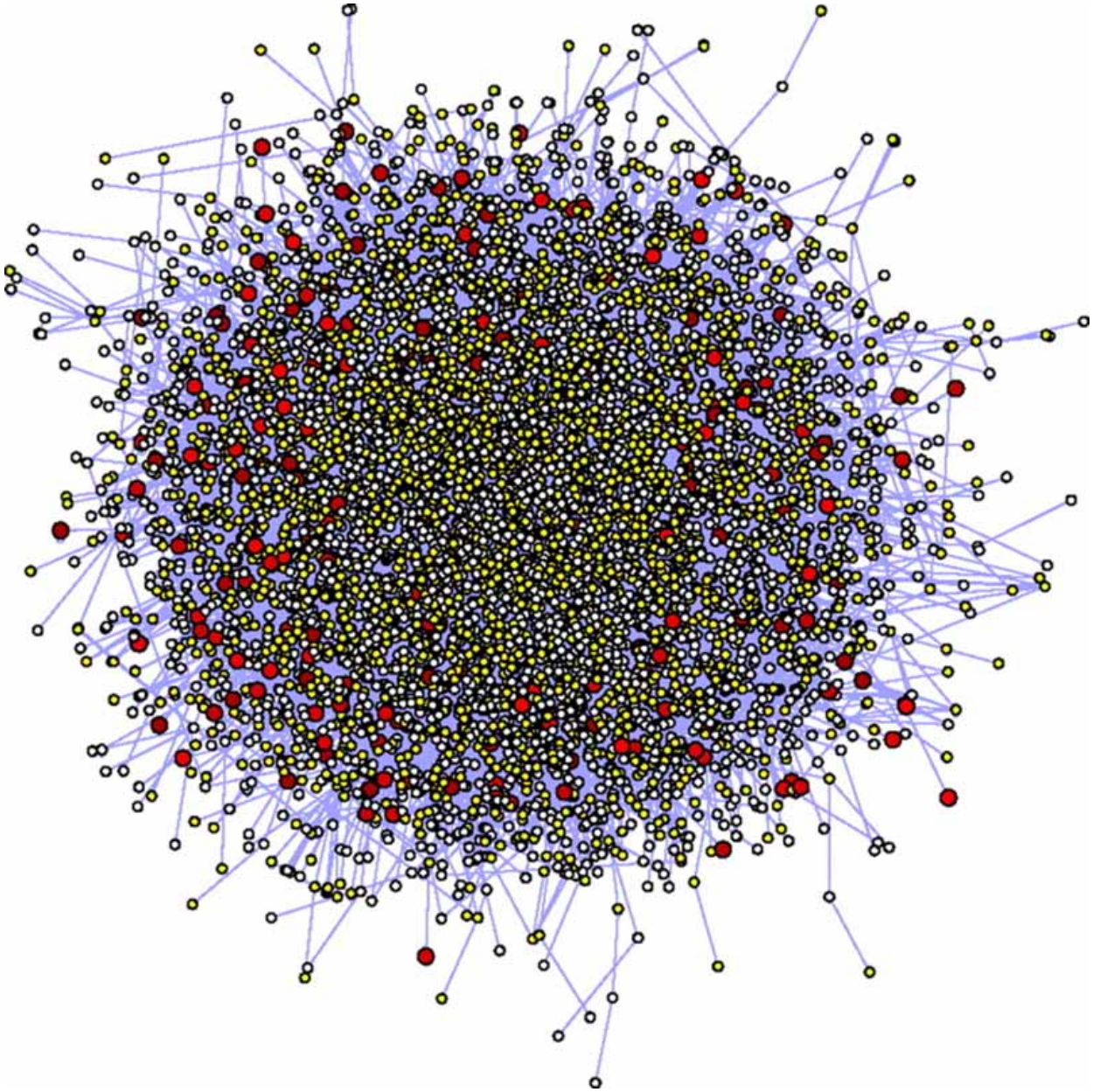
Barabási, A. L., Gulbahce, N. & Loscalzo, J. 2011. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics, 12*, ***56-68.***

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.

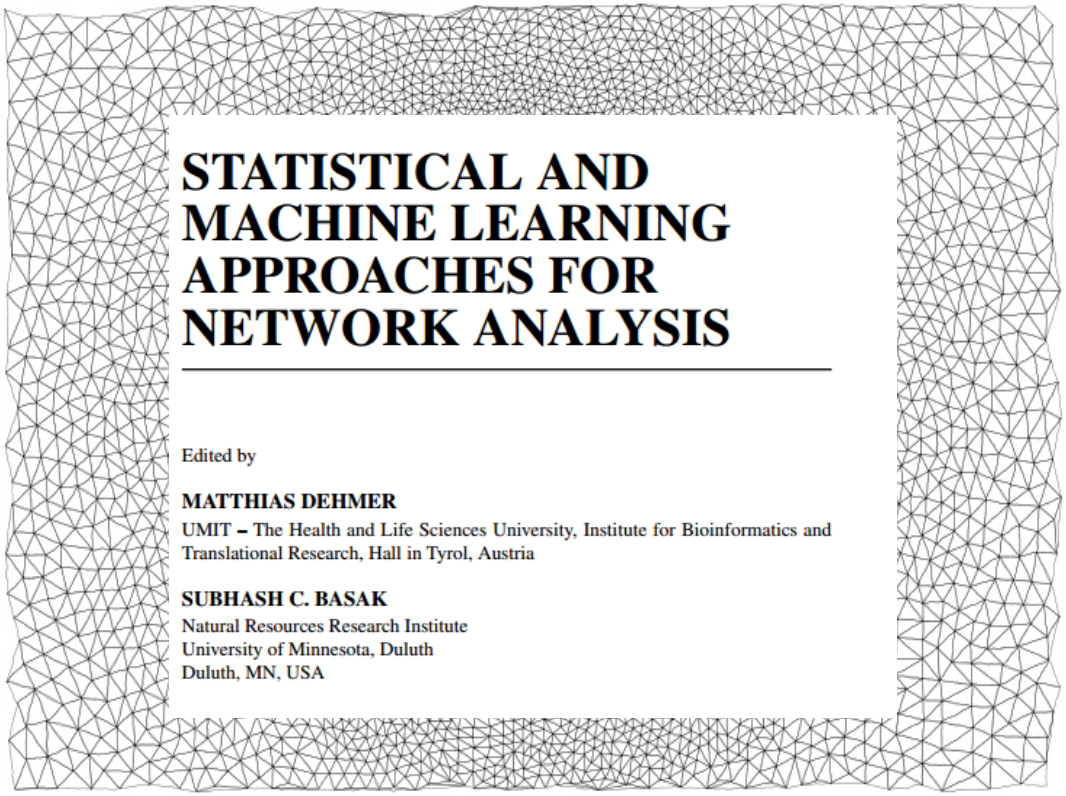
Example for a weakly structured data set - PPI

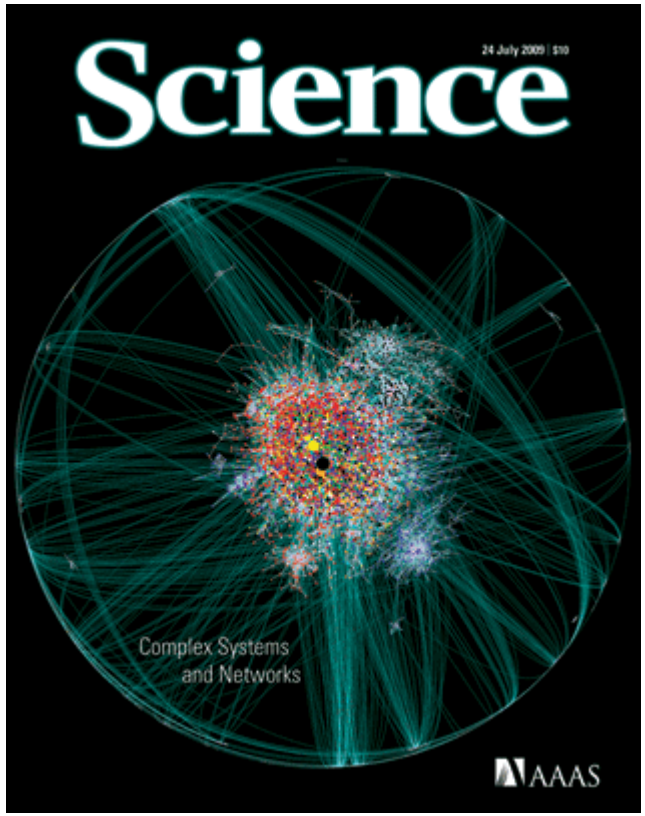


Kim, P. M., Korbel, J. O. & Gerstein, M. B. 2007. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. Proceedings of the National Academy of Sciences, 104, (51), 20274-20279.

4) 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

Network of Networks in Biology

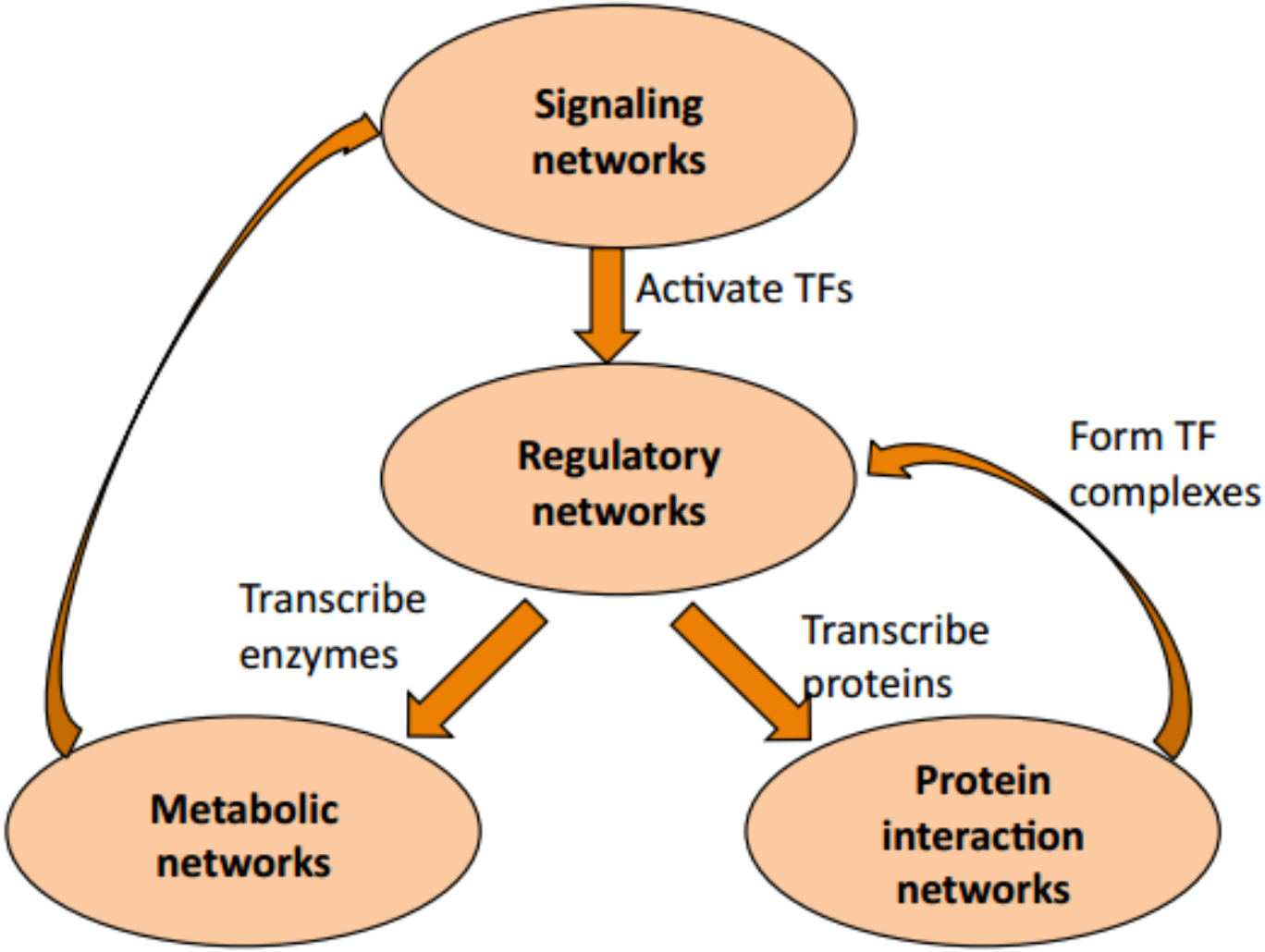


Image credit to Anna Goldenberg, Toronto

Genome-Phenome association in complex diseases

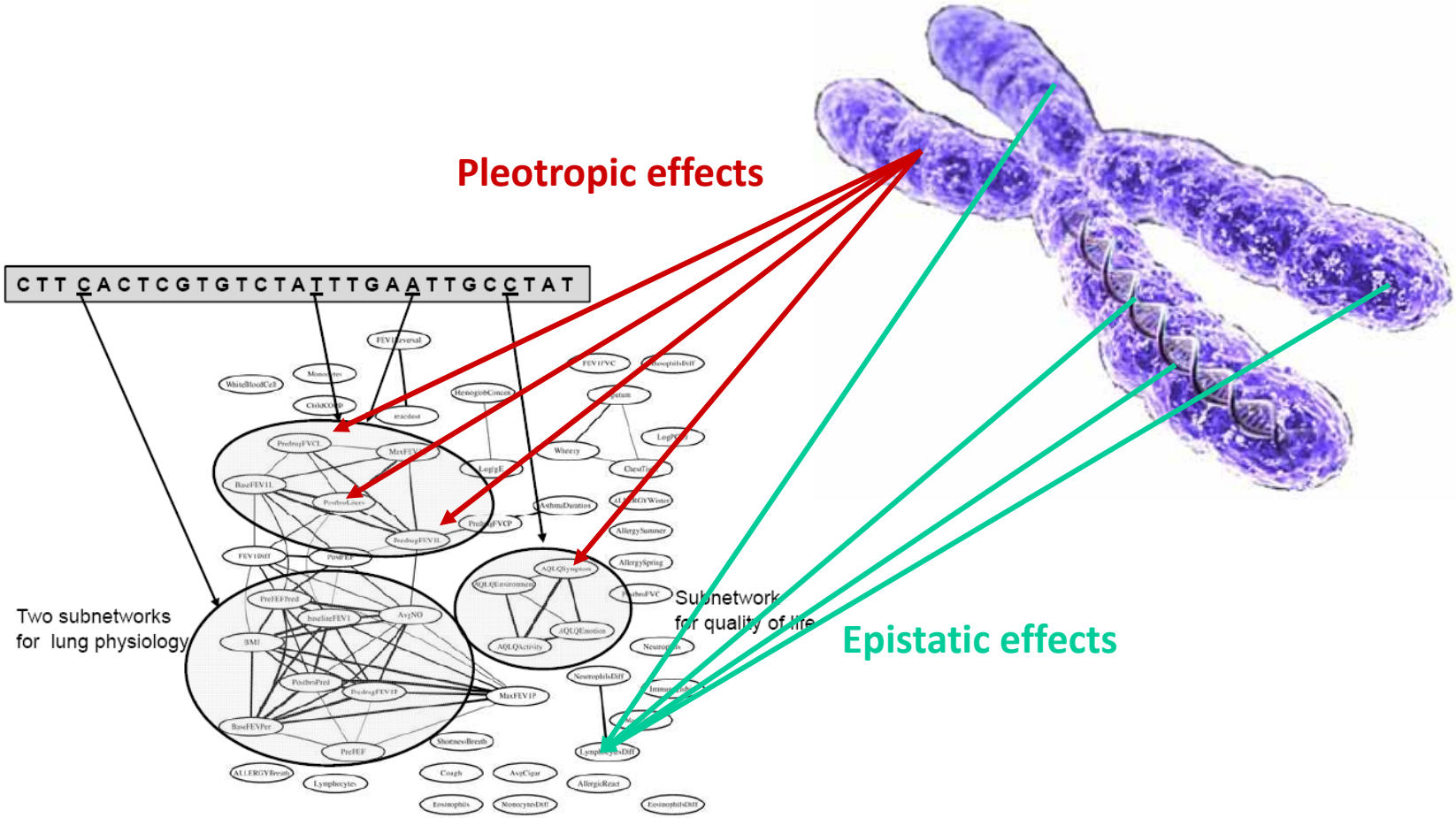
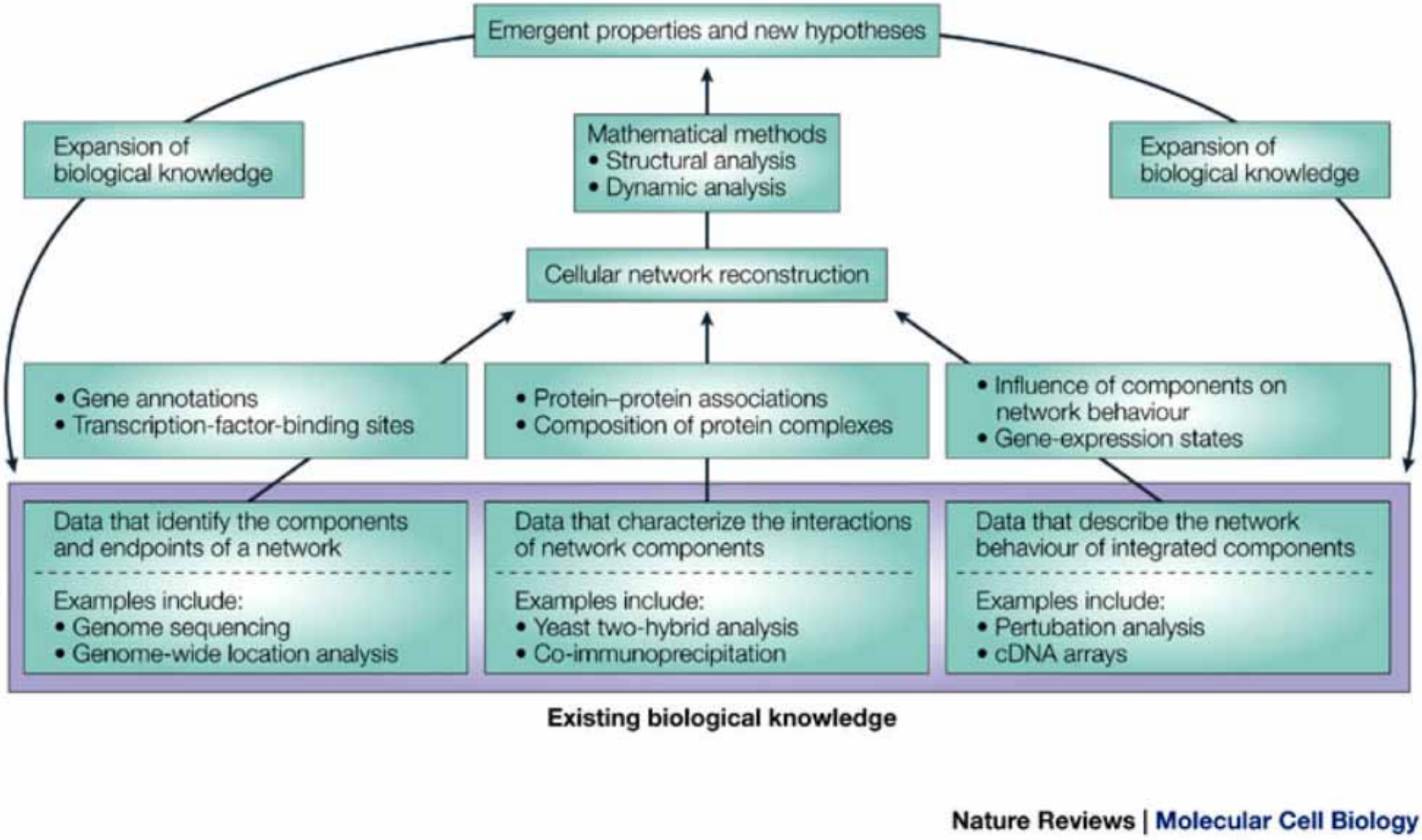


Image credit to Eric Xing, Carnegie Mellon University, Pittsburgh

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

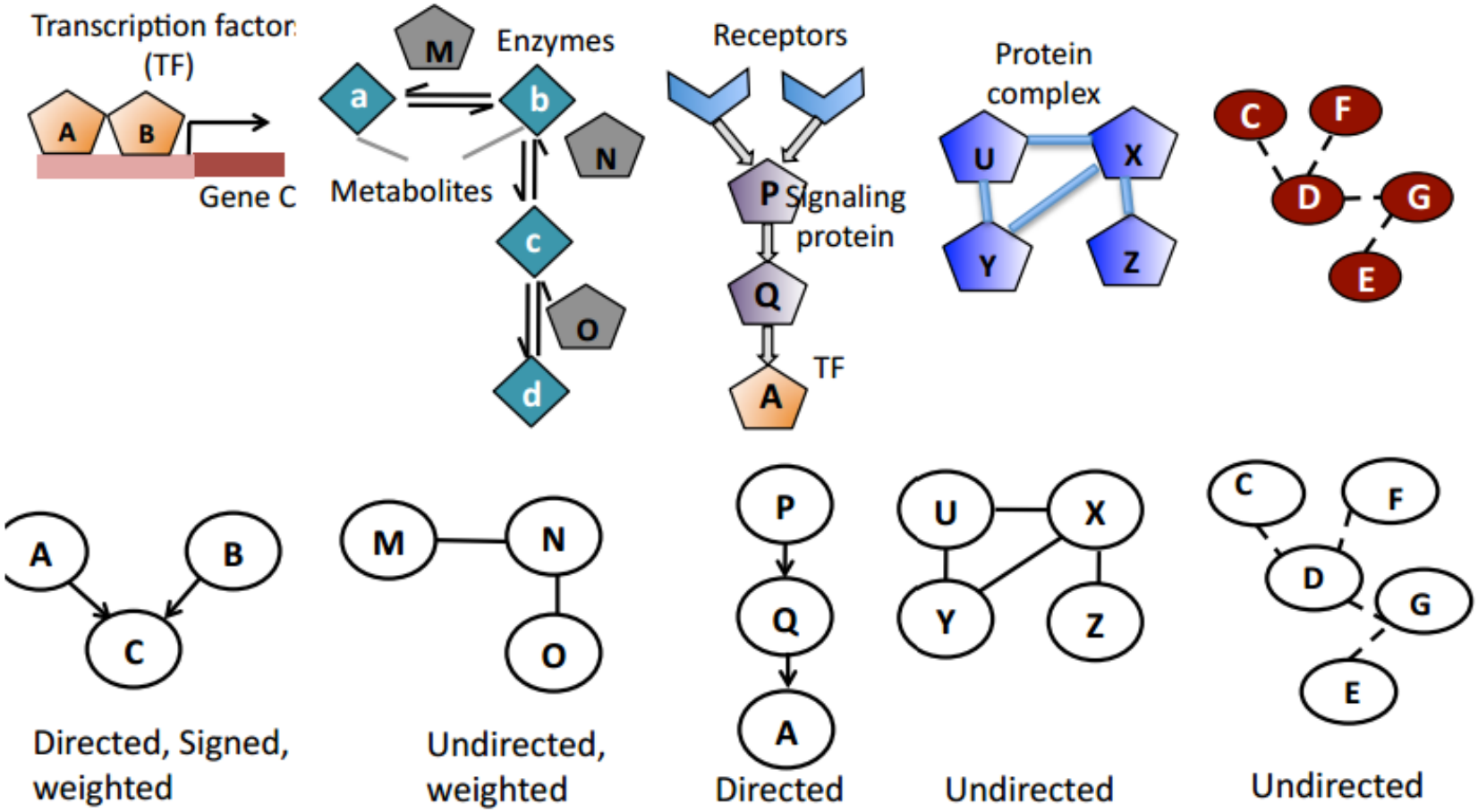
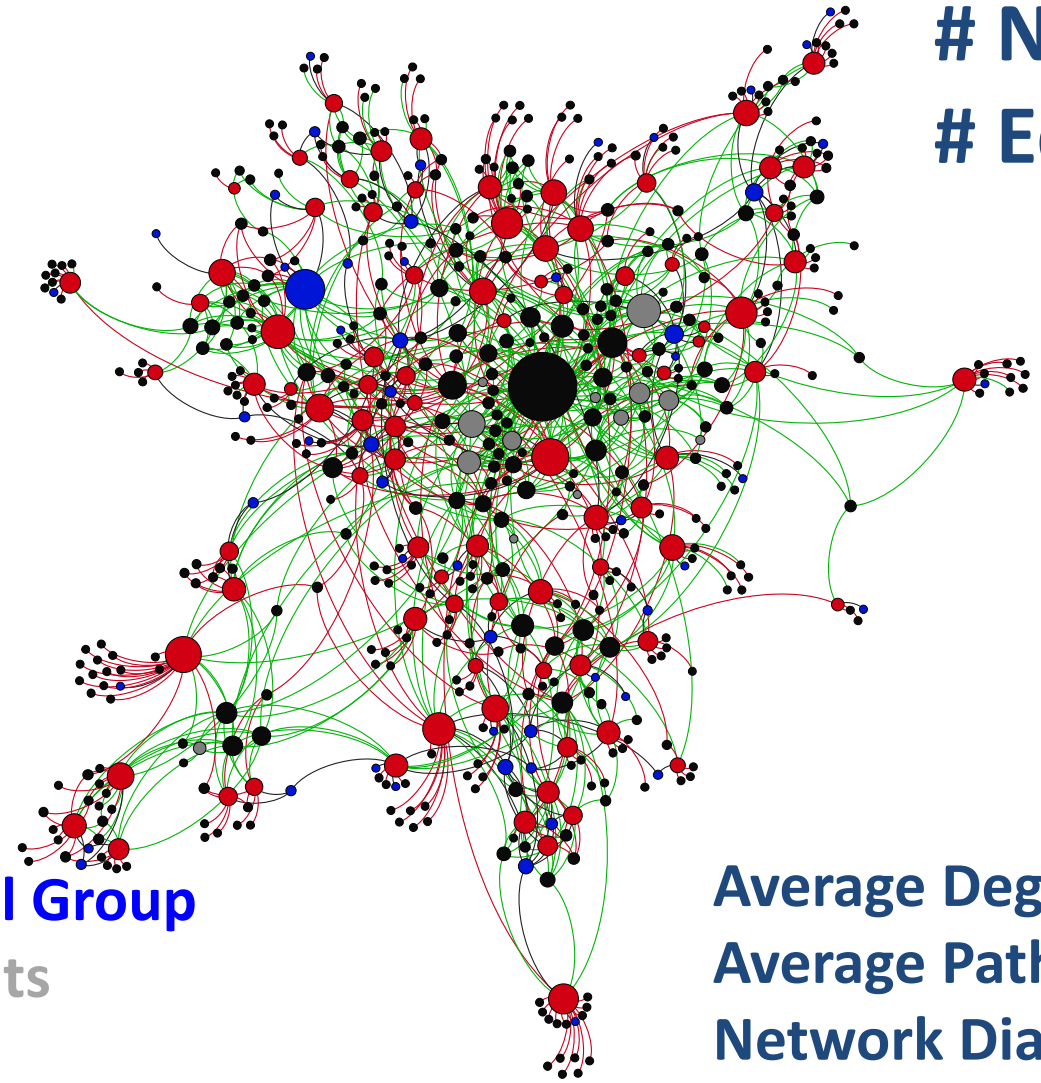


Image credit to Anna Goldenberg, Toronto

Example for a Medical Knowledge Space



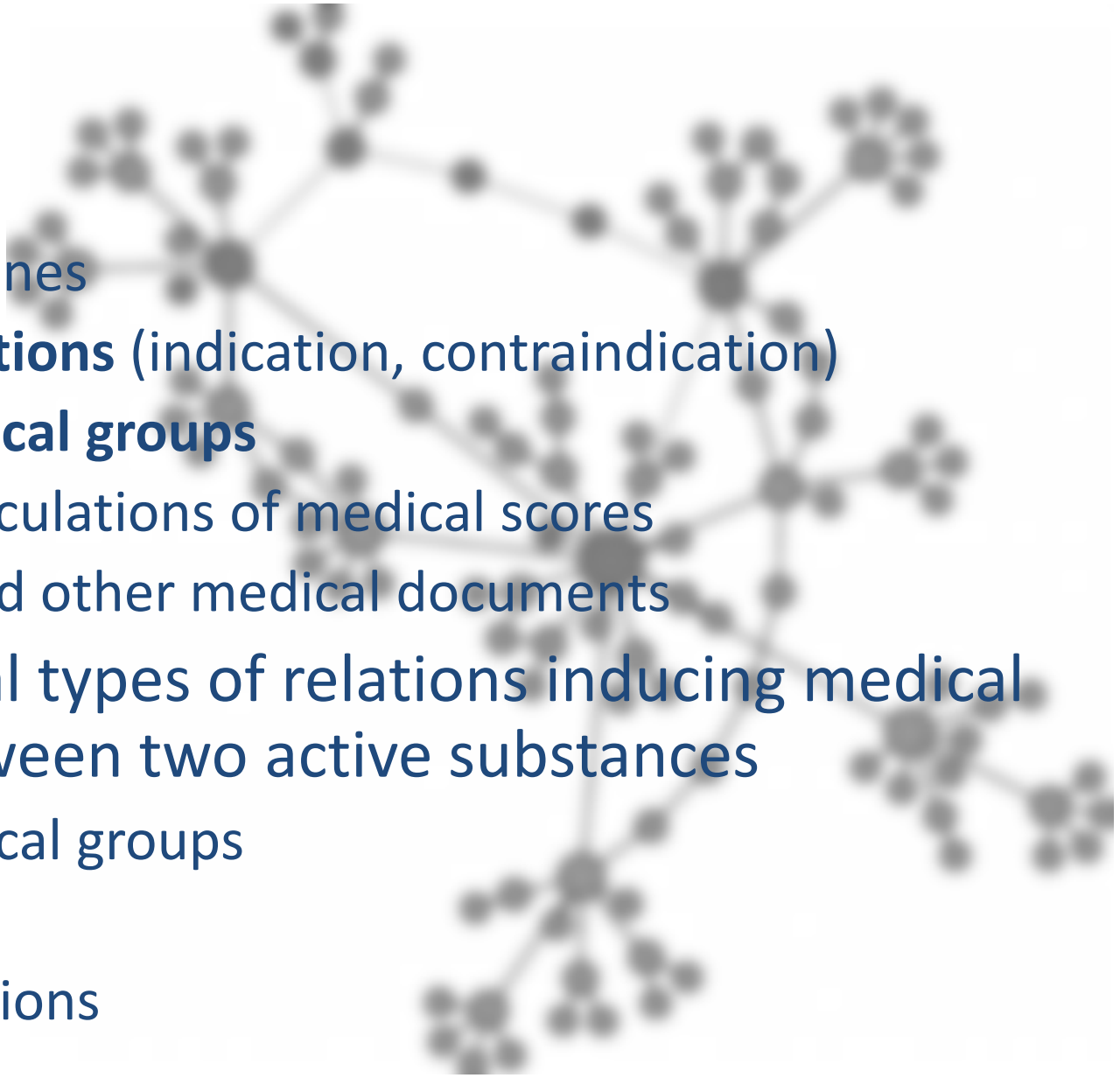
 # Nodes: 641
 # Edges: 1250

Agent
Condition
Pharmacological Group
Other Documents

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

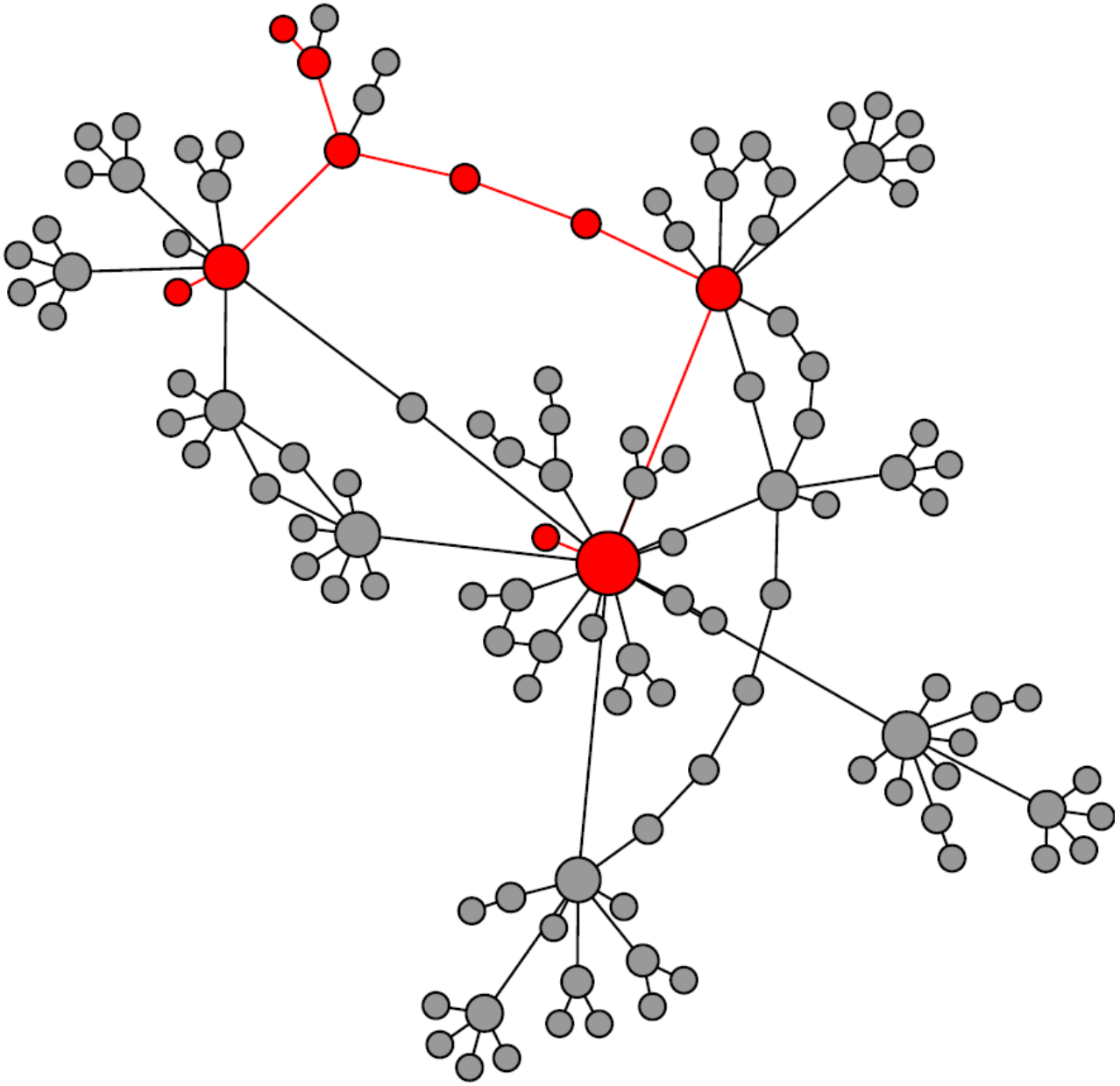
Holzinger, A., Ofner, B., Dehmer, M.: Multi-touch Graph-Based Interaction for Knowledge Discovery on Mobile Devices: State-of-the-Art and Future Challenges. In: LNCS 8401, pp. 241–254, (2014)

Medical Details of the Graph

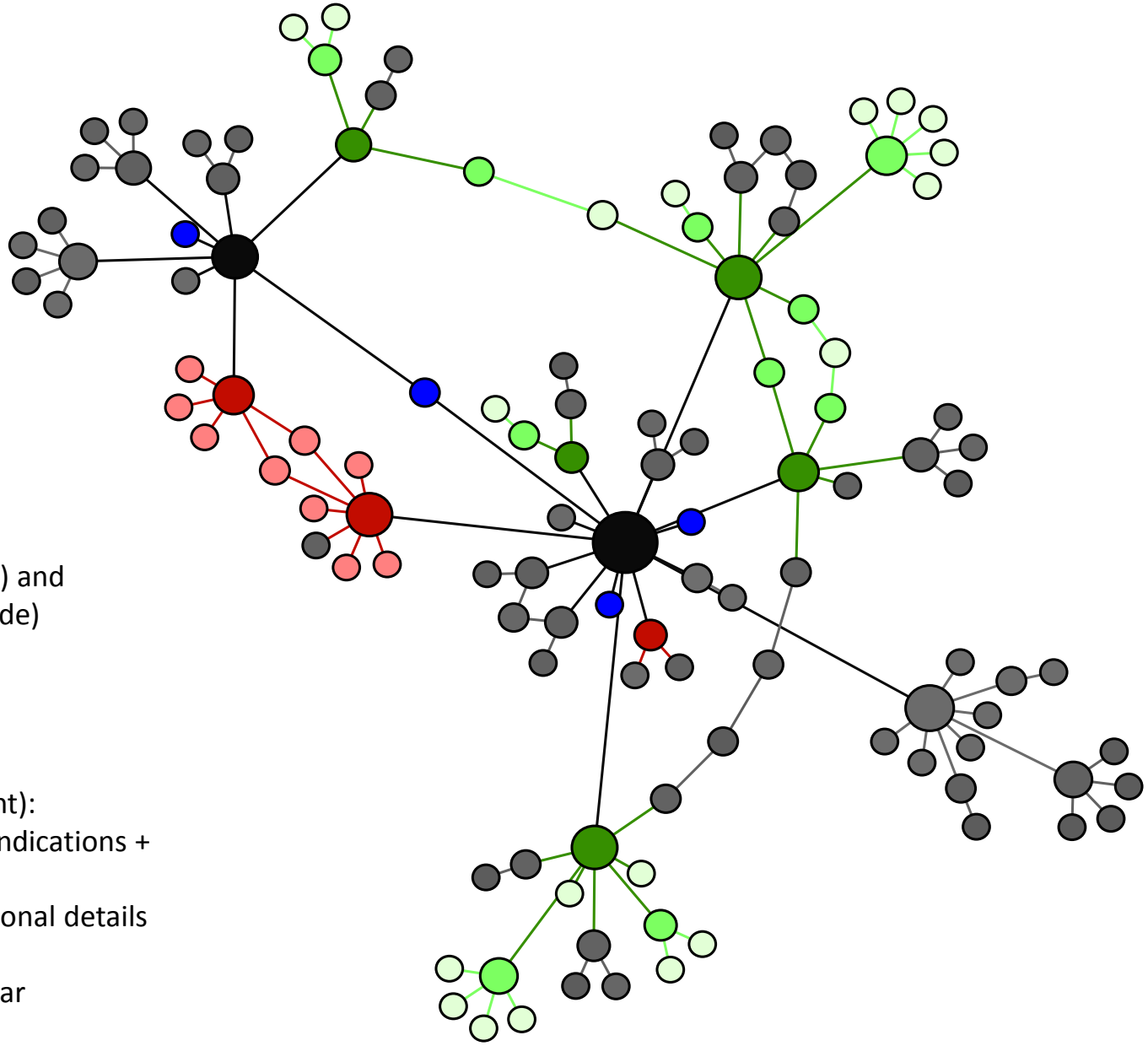


- **Nodes**
  - **drugs**
  - clinical guidelines
  - **patient conditions** (indication, contraindication)
  - **pharmacological groups**
  - tables and calculations of medical scores
  - algorithms and other medical documents
- **Edges**: 3 crucial types of relations inducing medical relevance between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

Example for the shortest path



Example for finding related structures

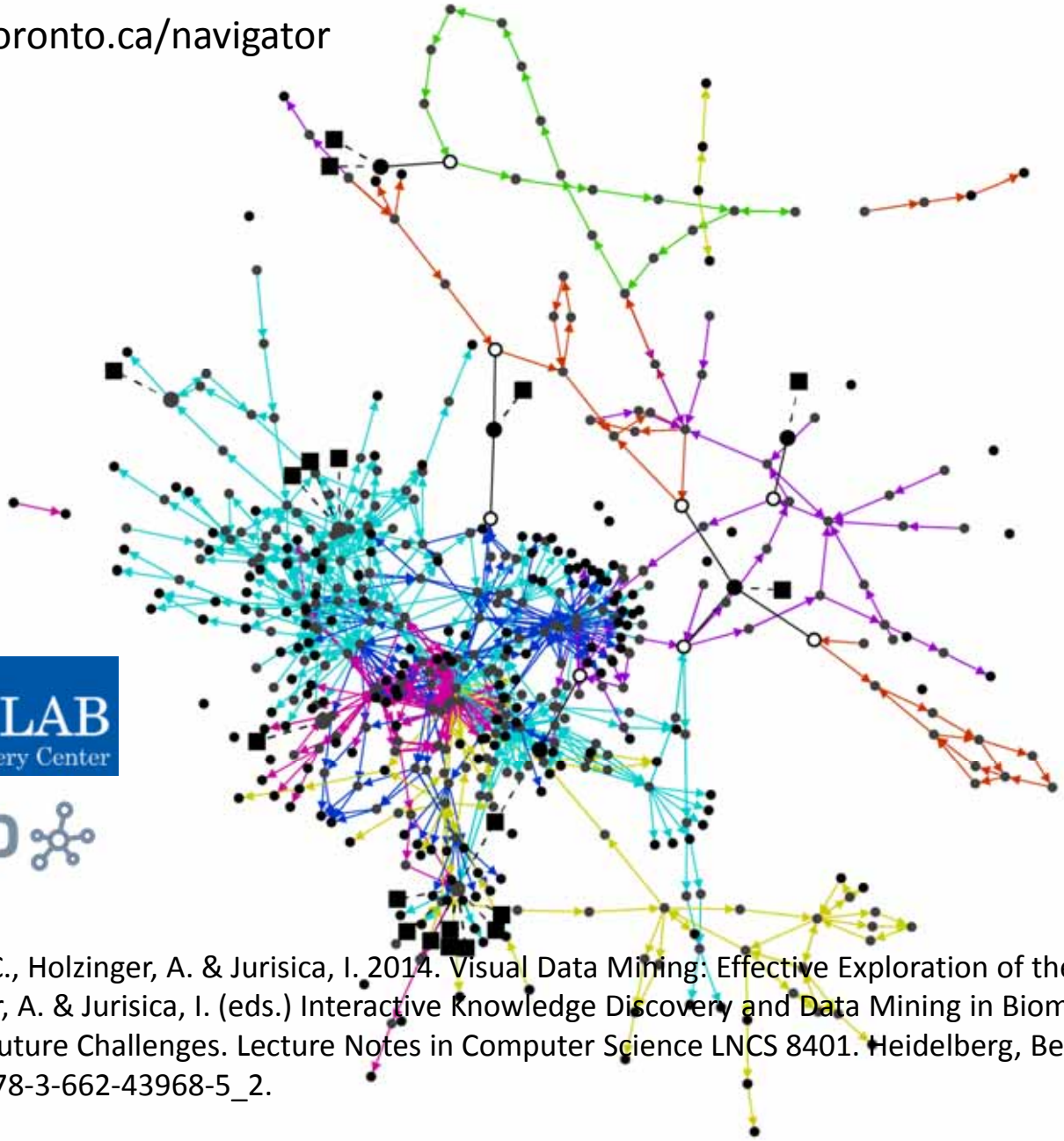


Relationship between
Adrenaline (center black node) and
Dobutamine (top left black node)
Blue: Pharmacological Group
Dark red: Contraindication;
Light red: Condition

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

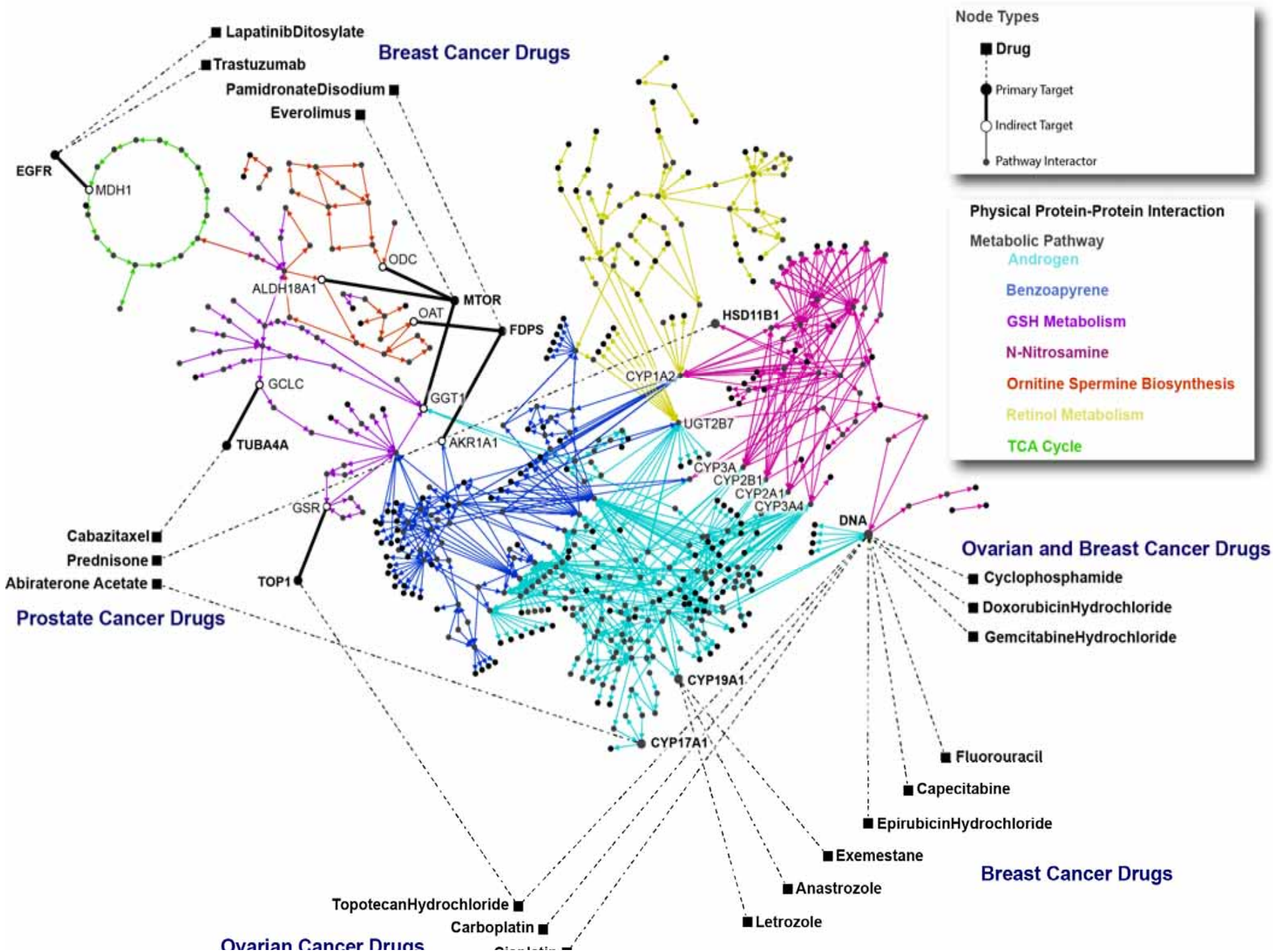
http://ophid.utoronto.ca/navigator



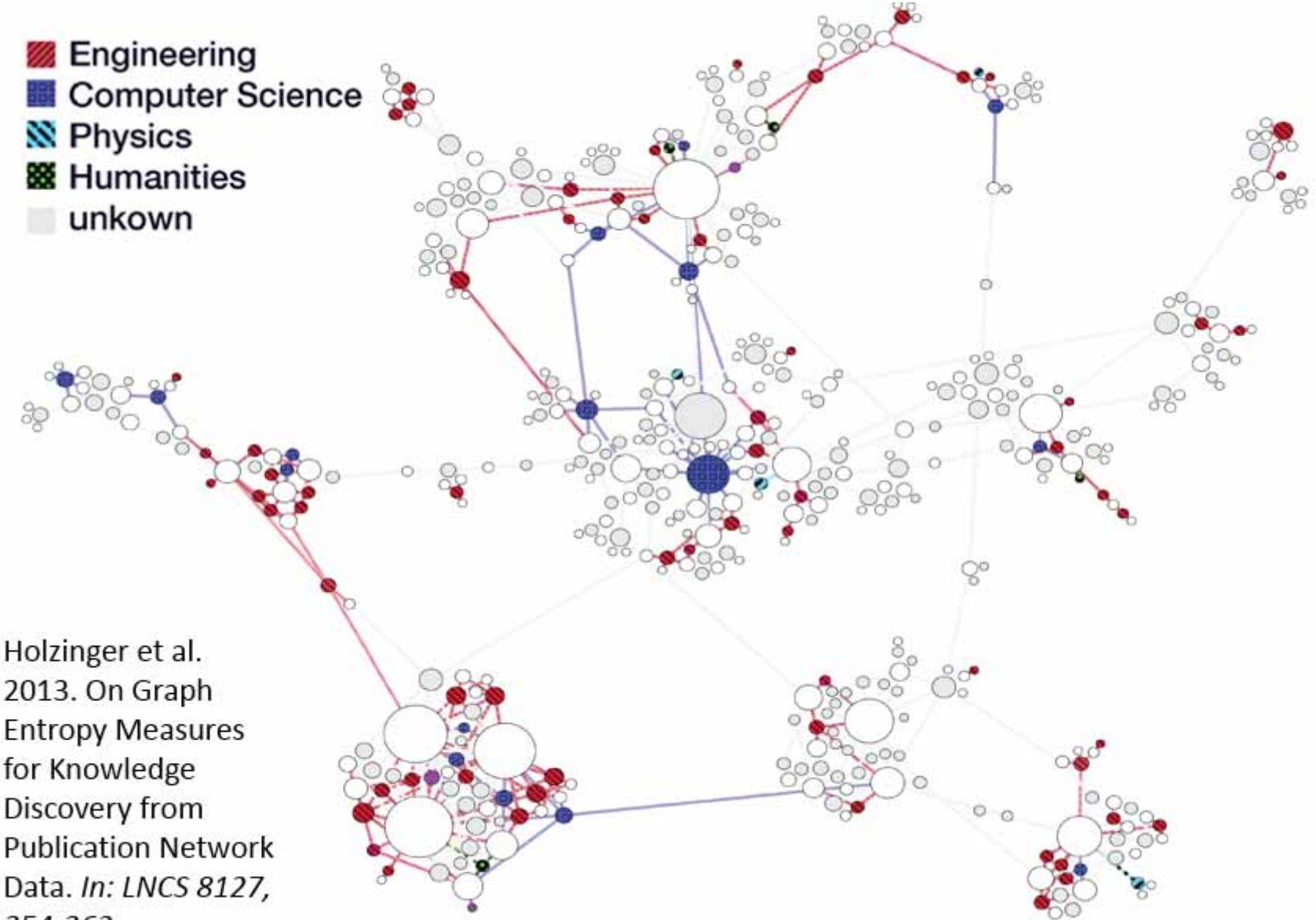
JURISICA LAB
IBM Life Sciences Discovery Center

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.


Node Types
Drug
Primary Target
Indirect Target
Pathway Interactor
Physical Protein-Protein Interaction
Metabolic Pathway
Androgen
Benzoapyrene
GSH Metabolism
N-Nitrosamine
Ornitine Spermine Biosynthesis
Retinol Metabolism
TCA Cycle
Breast Cancer Drugs
LapatinibDitosylate
Trastuzumab
PamidronateDisodium
Everolimus
EGFR
MDH1
ODC
ALDH18A1
MTOR
OAT
FDPS
GCLC
GGT1
AKR1A1
TUBA4A
HSD11B1
CYP1A2
UGT2B7
CYP3A
CYP2B1
CYP2A1
CYP3A4
GSR
DNA
Cabazitaxel
Prednisone
Abiraterone Acetate
TOP1
Prostate Cancer Drugs
Ovarian and Breast Cancer Drugs
Cyclophosphamide
DoxorubicinHydrochloride
GemcitabineHydrochloride
CYP19A1
CYP17A1
Fluorouracil
Capecitabine
EpirubicinHydrochloride
Exemestane
Anastrozole
Letrozole
Breast Cancer Drugs
TopotecanHydrochloride
Carboplatin
Ovarian Cancer Drugs

Example: Graph Entropy Measures



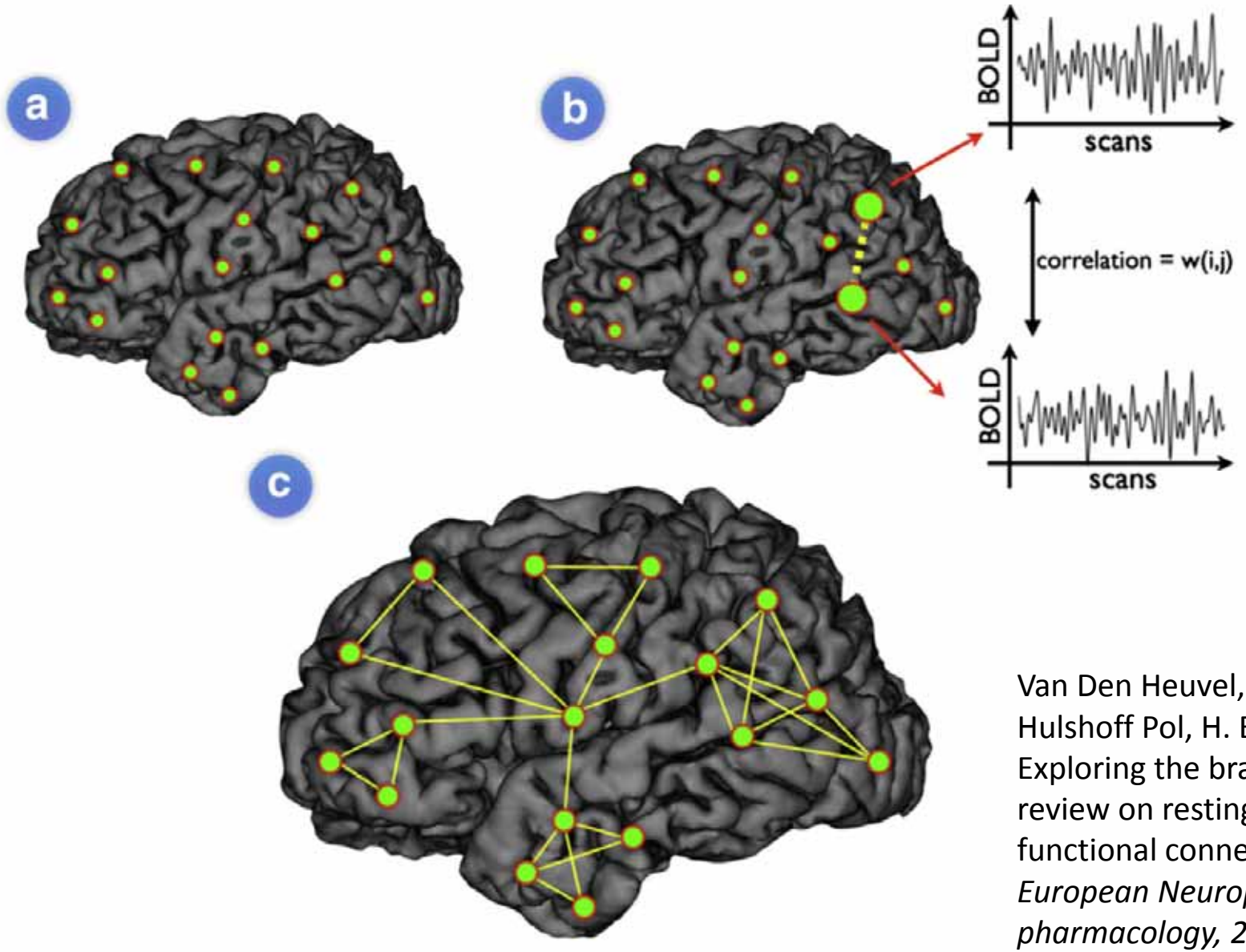
Holzinger et al. 2013. On Graph Entropy Measures for Knowledge Discovery from Publication Network Data. *In: LNCS 8127, 354-362.*

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.

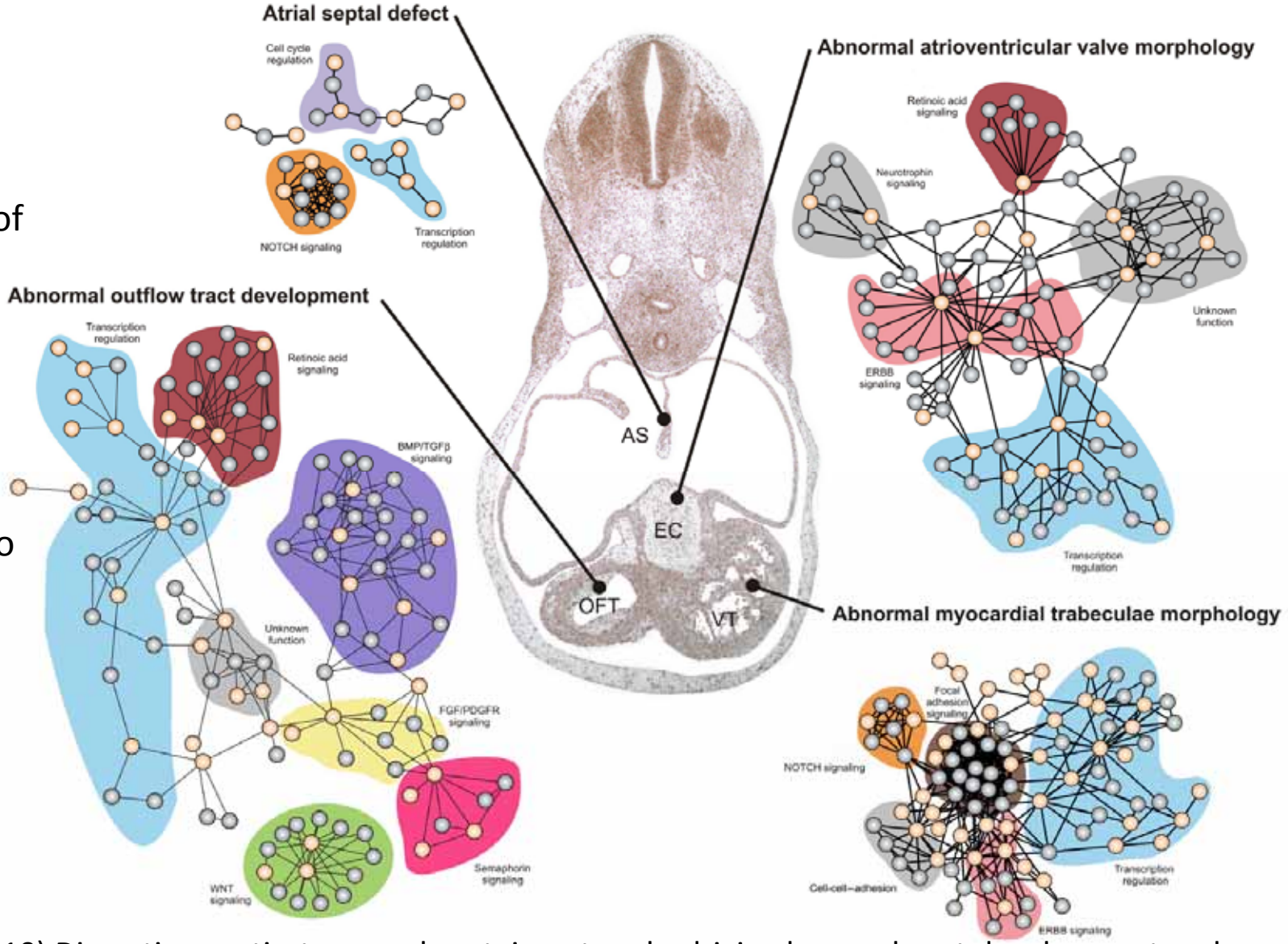
Example: The brain is a complex network



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

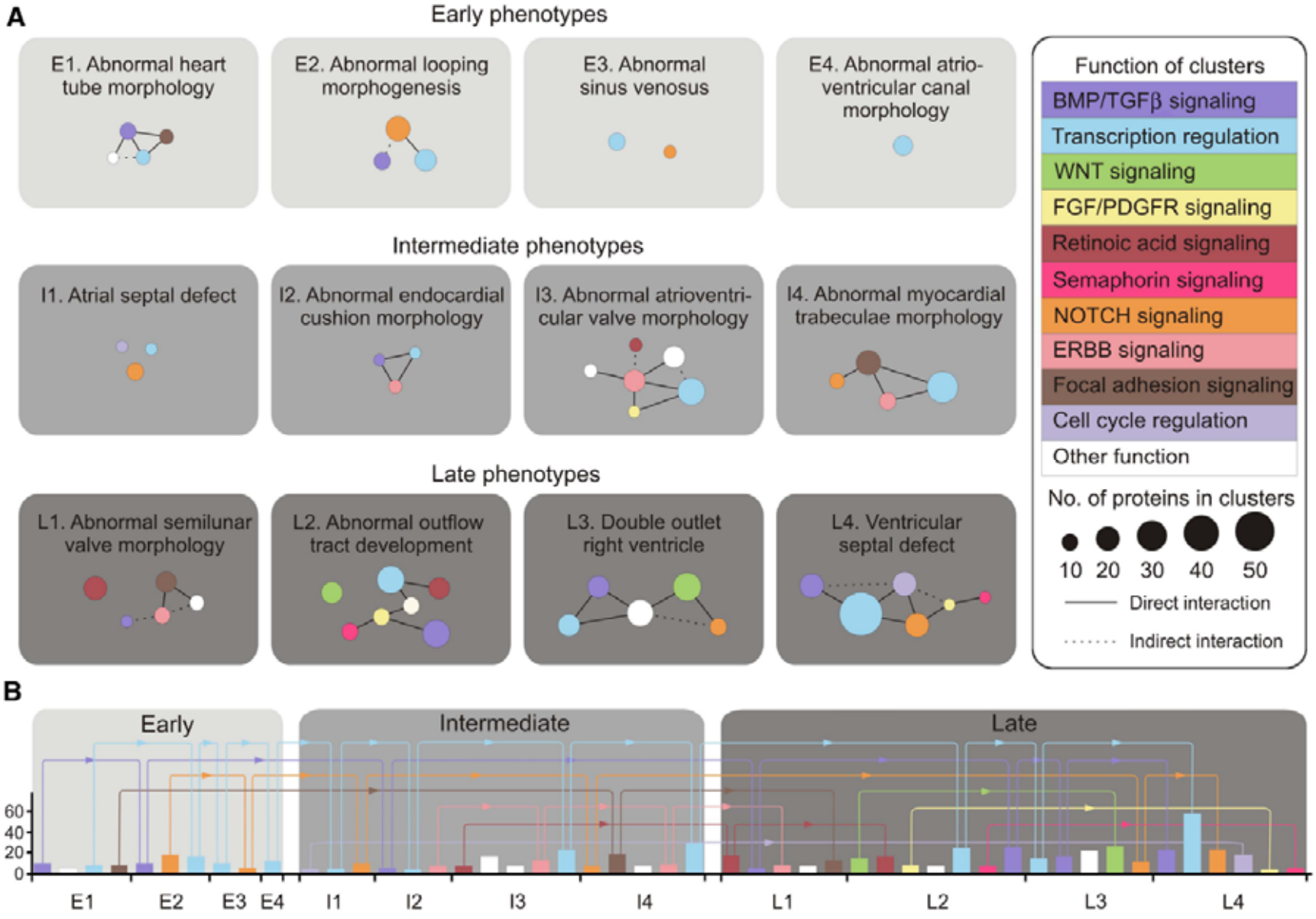
Representative Examples of disease complexes

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



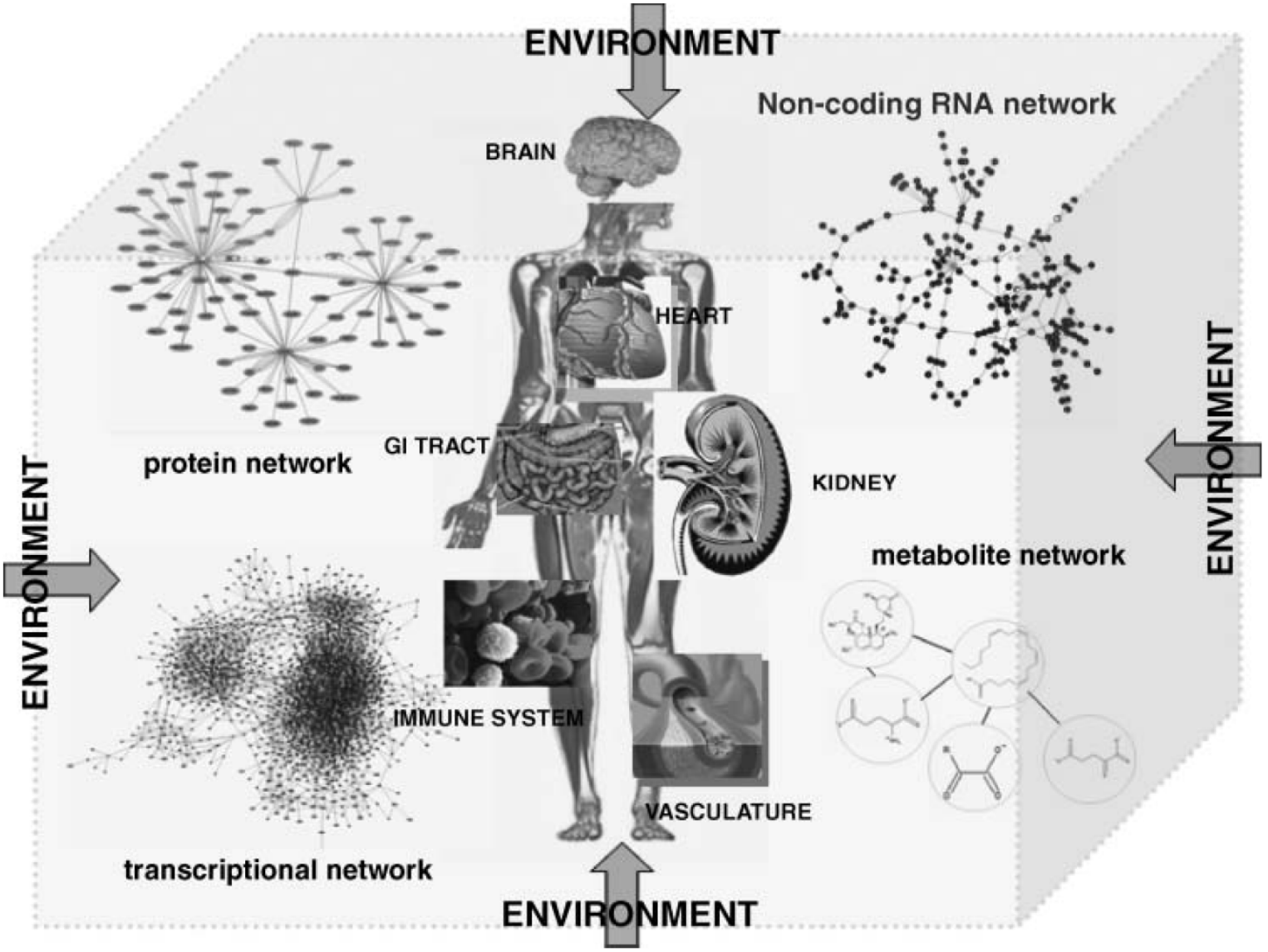
Lage, K. et. al (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular systems biology, 6, 1, 1-9*.

Example: Cell-based therapy



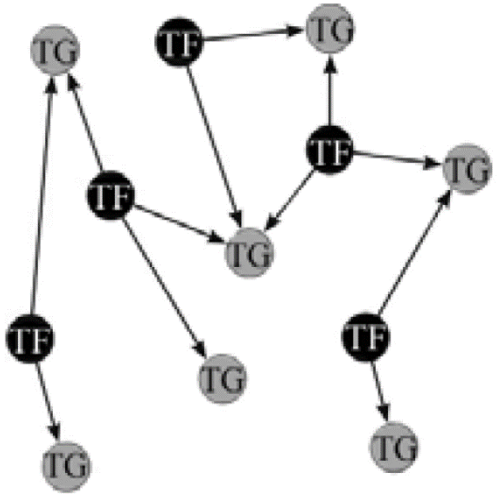
Lage et. al (2010)

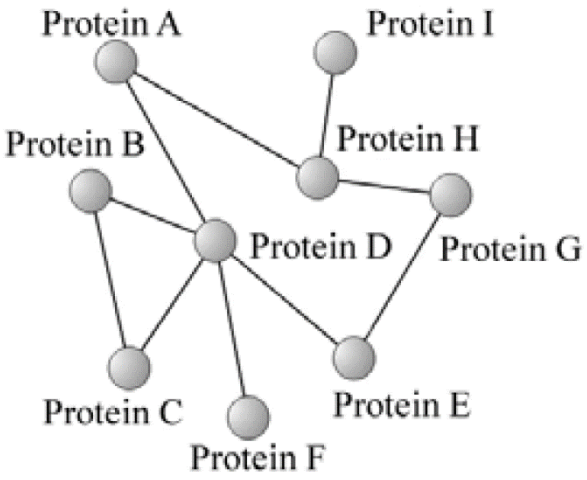
Identifying Networks in Disease Research

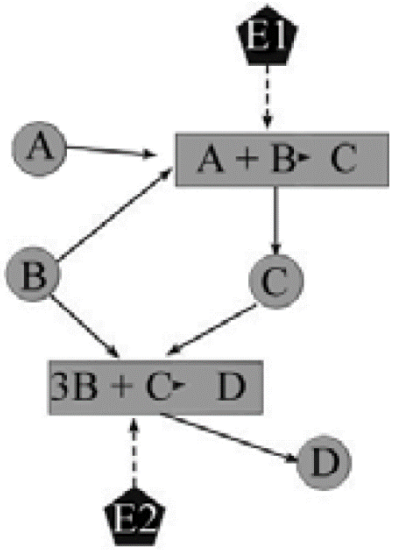


Schadt, E. E. & Lum, P. Y. (2006) Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *Journal of lipid research, 47, 12, 2601-2613.*

Three main types of biomedical networks







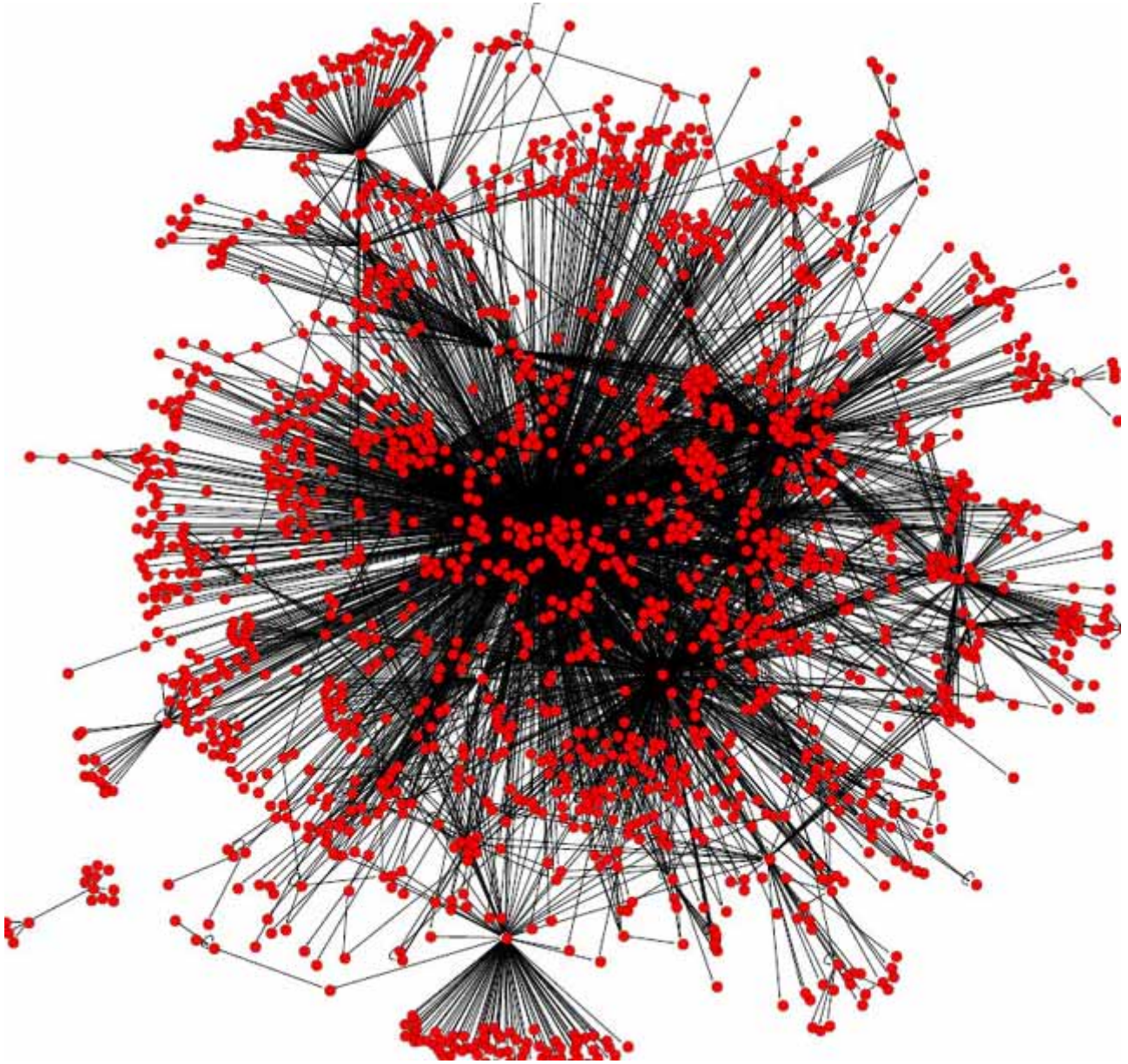
Transcriptional regulatory network with two components:
TF = transcription factor
TG = target genes
(TF regulates the transcription of TG)

Protein-Protein interaction network

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

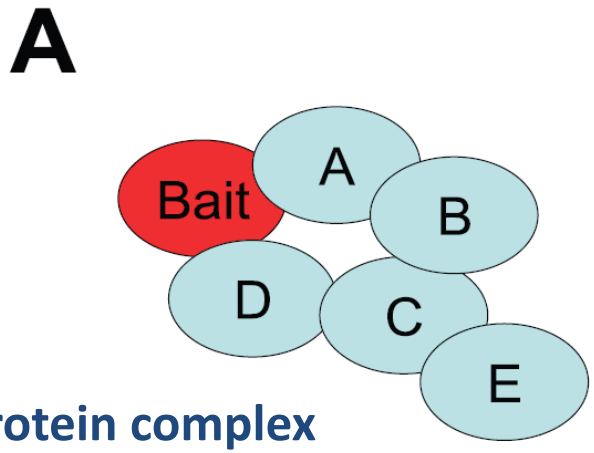
Costa, L. F., Rodrigues, F. A. & Cristino, A. S. (2008) Complex networks: the key to systems biology. *Genetics and Molecular Biology, 31, 3, 591–601.*

Example Transcriptional Regulatory Network

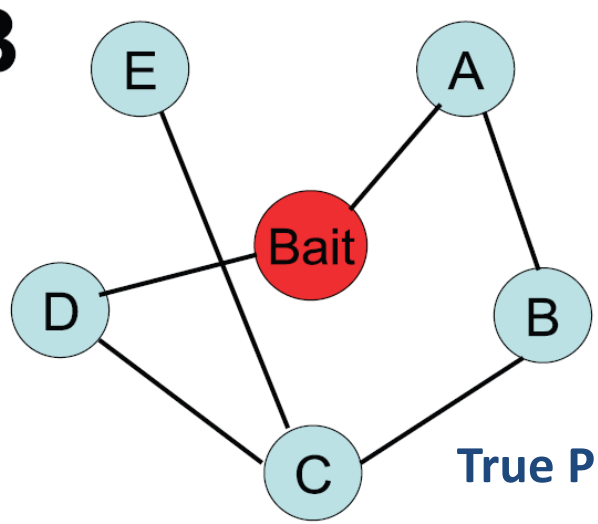


Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Peralta-Gil, M., Peñaloza-Spínola, M. I., Martínez-Antonio, A., Karp, P. D. & Collado-Vides, J. 2006. The comprehensive updated regulatory network of Escherichia coli K-12. *BMC bioinformatics, 7, (1), 5.*

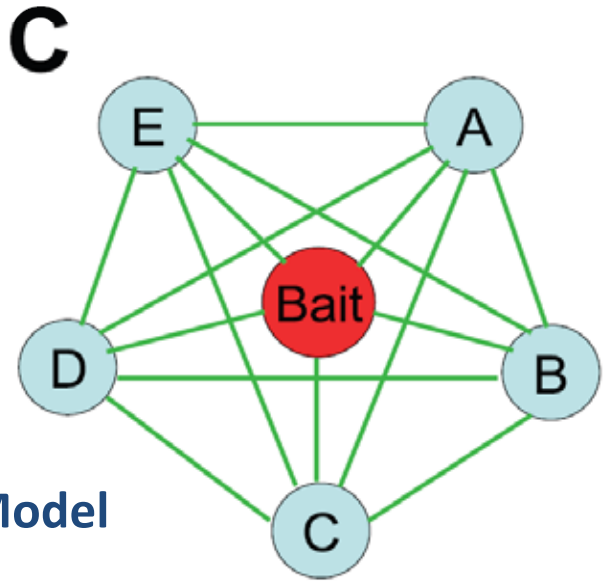
Network Representations of Protein Complexes



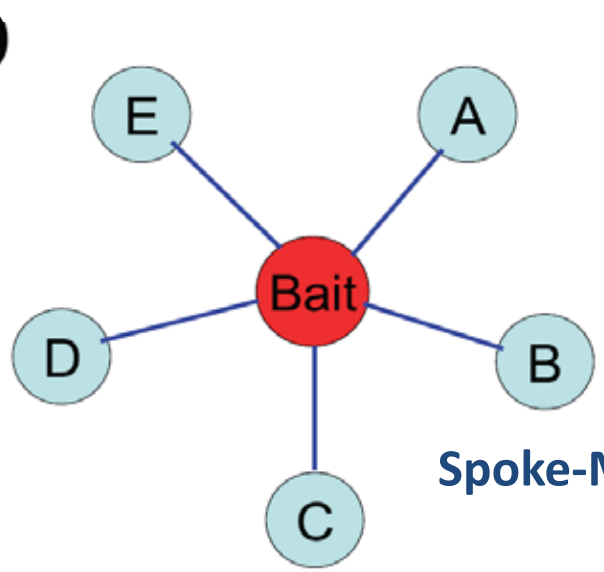
Protein complex



True PPI topology



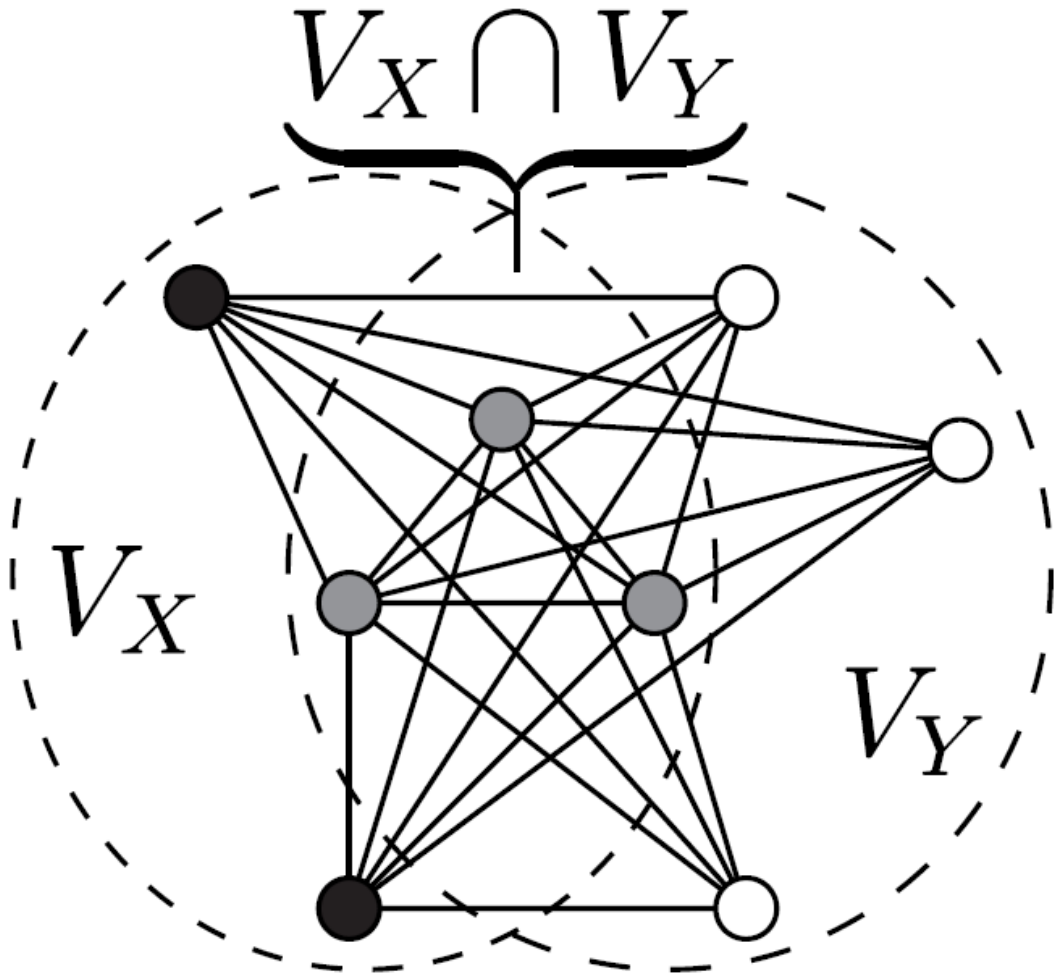
Matrix-Model



Spoke-Model

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)



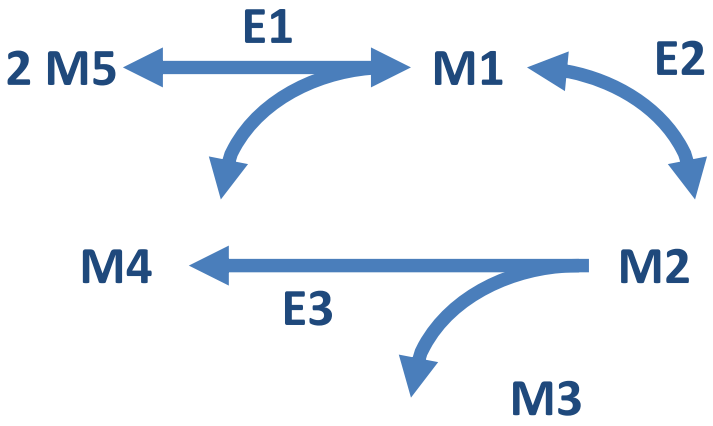
Boyen, P., Van Dyck, D., Neven, F., van Ham, R. C. H. J. & van Dijk, A. (2011) SLIDER: A Generic Metaheuristic for the Discovery of Correlated Motifs in Protein-Protein Interaction Networks. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on, 8, 5, 1344-1357.*

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

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

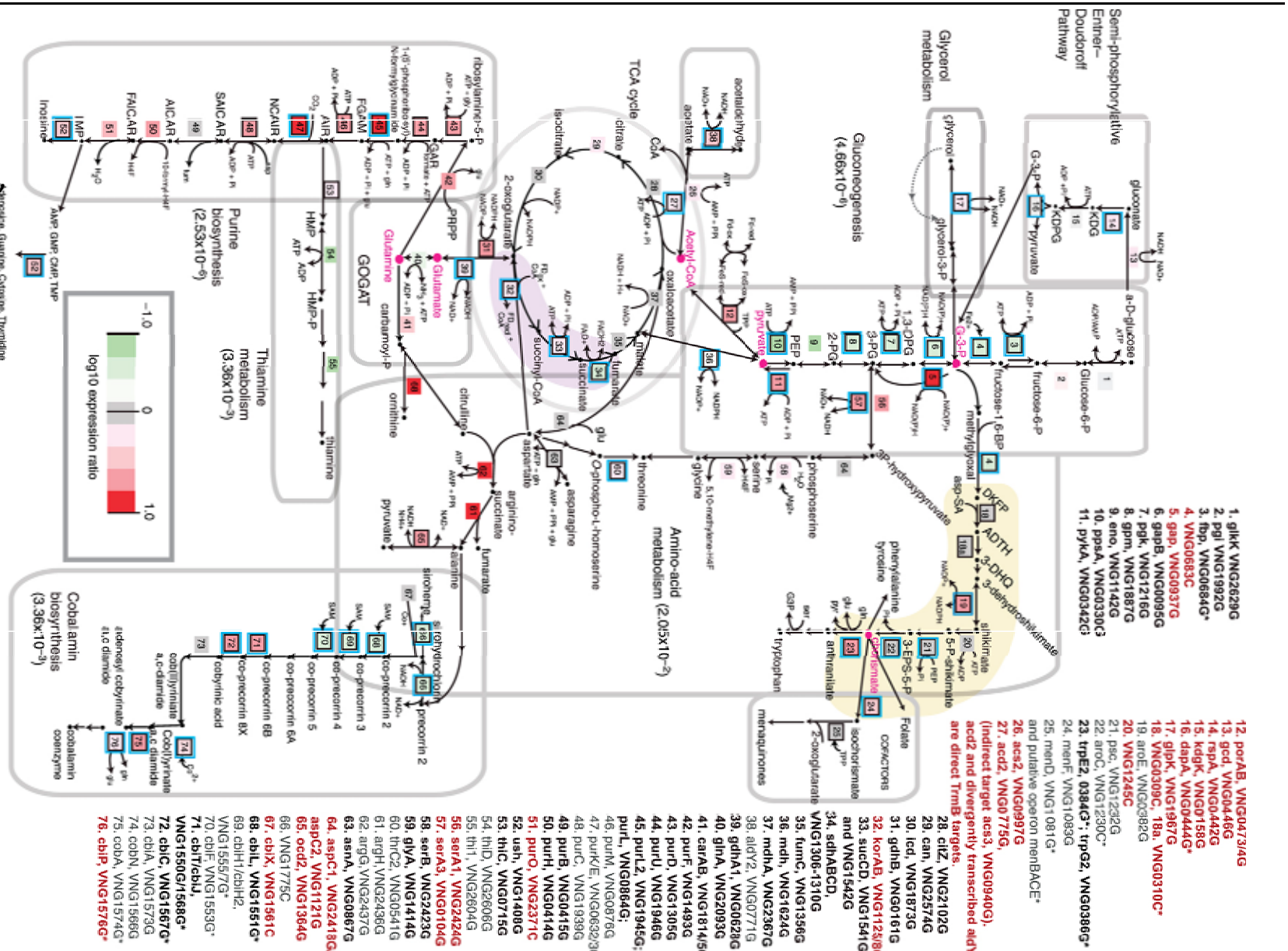
M1	M2
M1	M4
M1	M5
M2	M1
M2	M3
M2	M4
M4	M1
M5	M1

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 … big data ☺

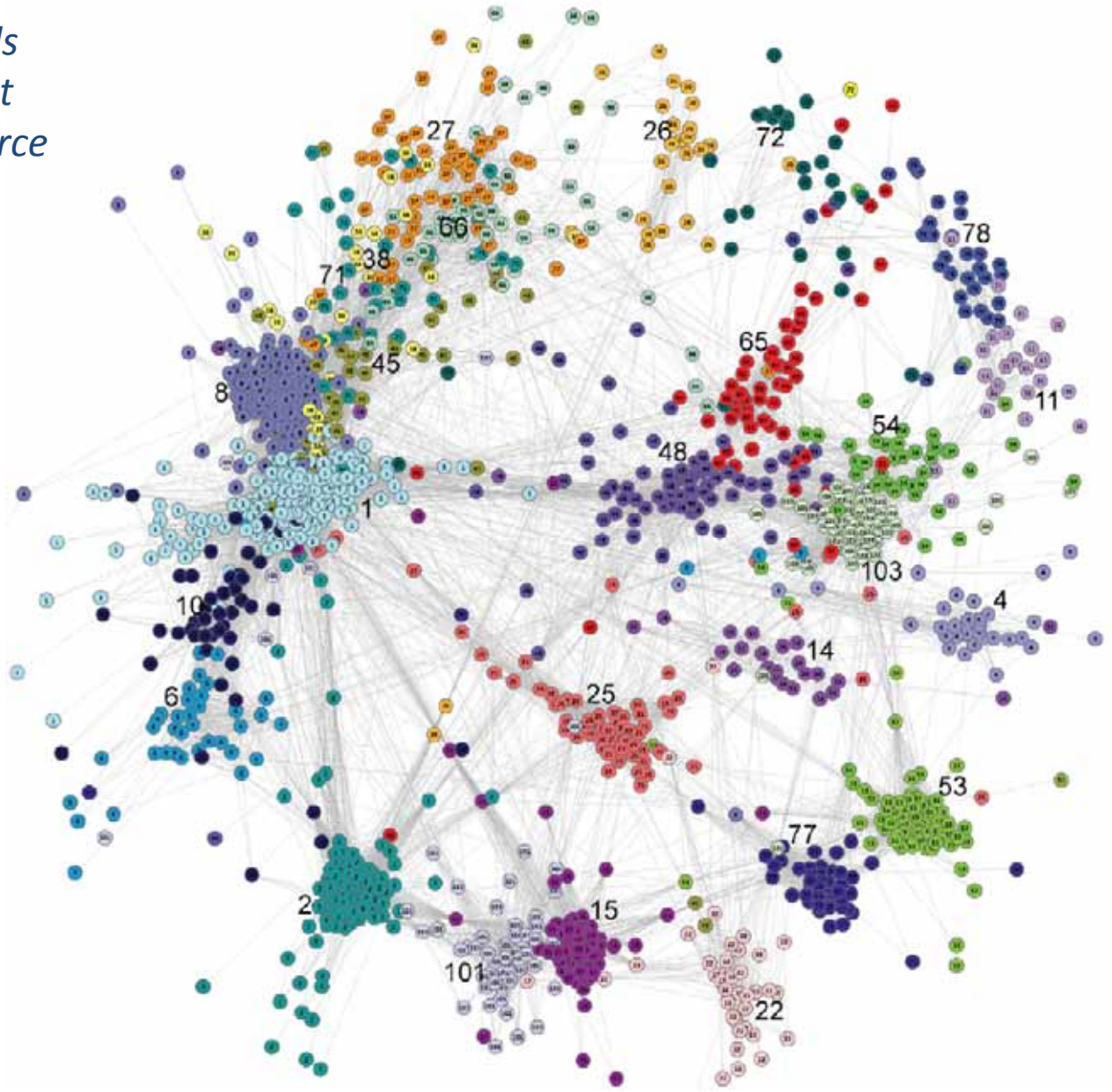
Schmid, A. K., Reiss, D. J., Pan, M., Koide, T. & Baliga, N. S. (2009) A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. *Molecular Systems Biology, 5, 1-9.*



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

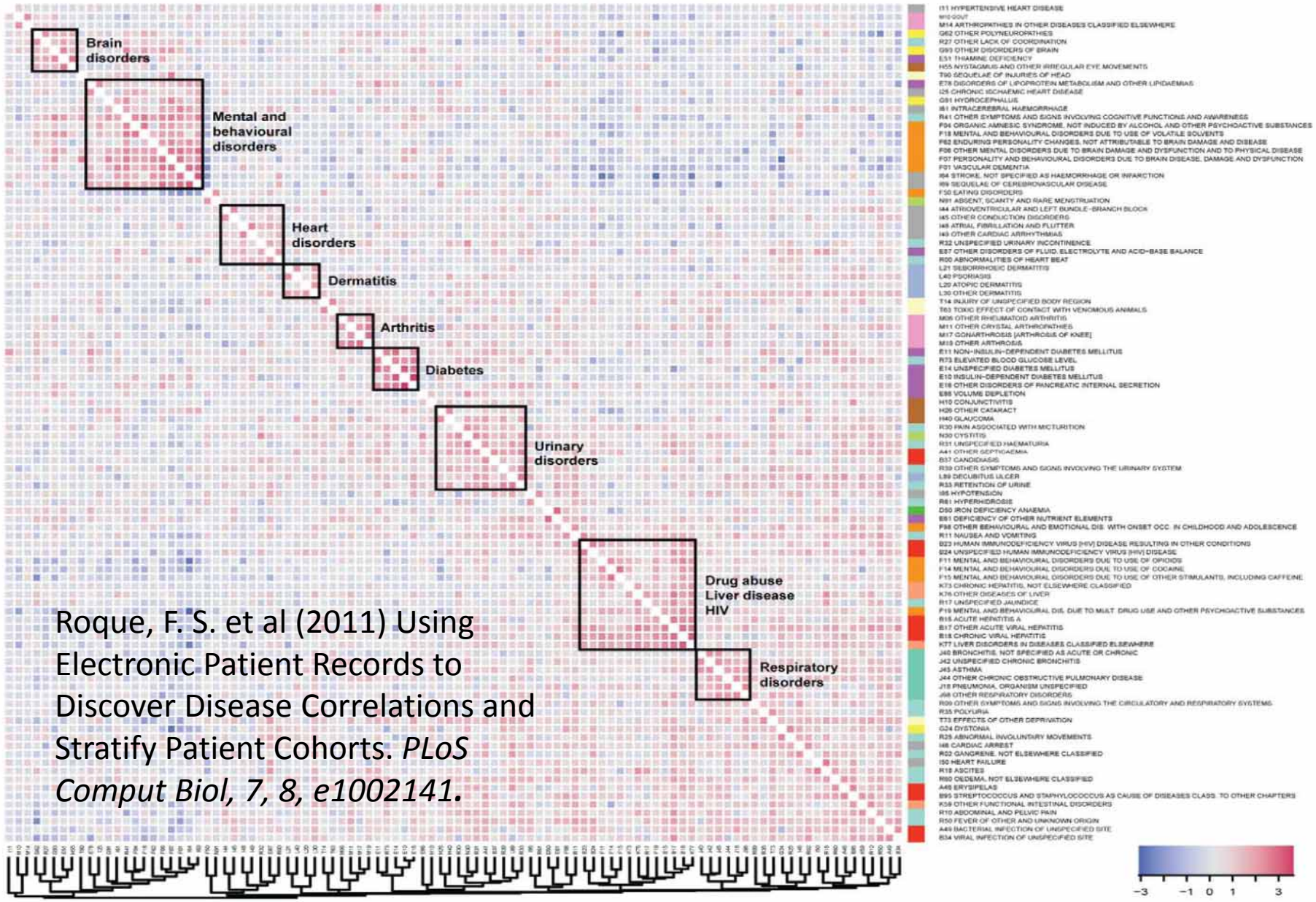
Using EPRs to Discover Disease Correlations

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.*

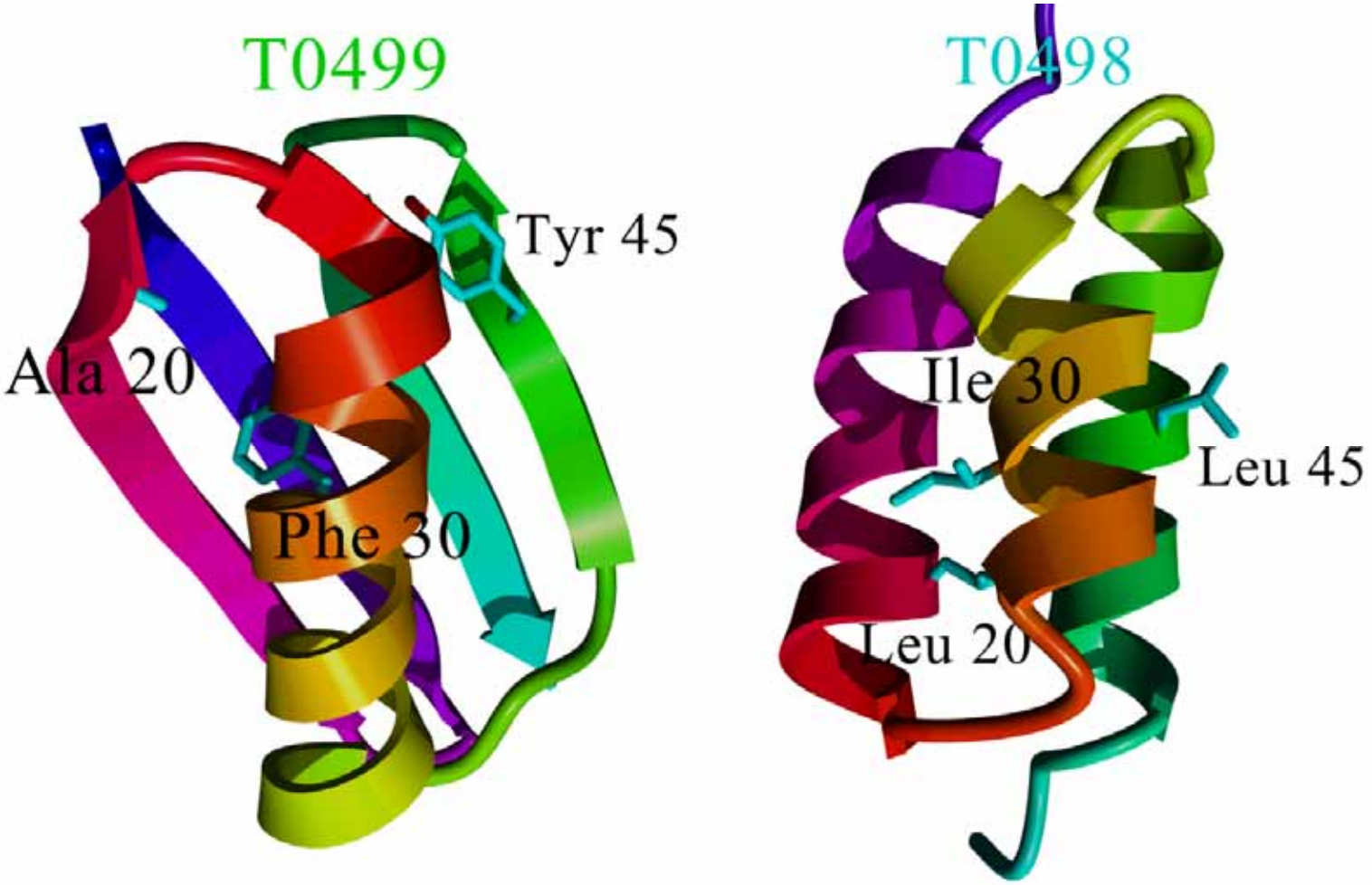
Heatmap of disease-disease correlations (ICD)



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)

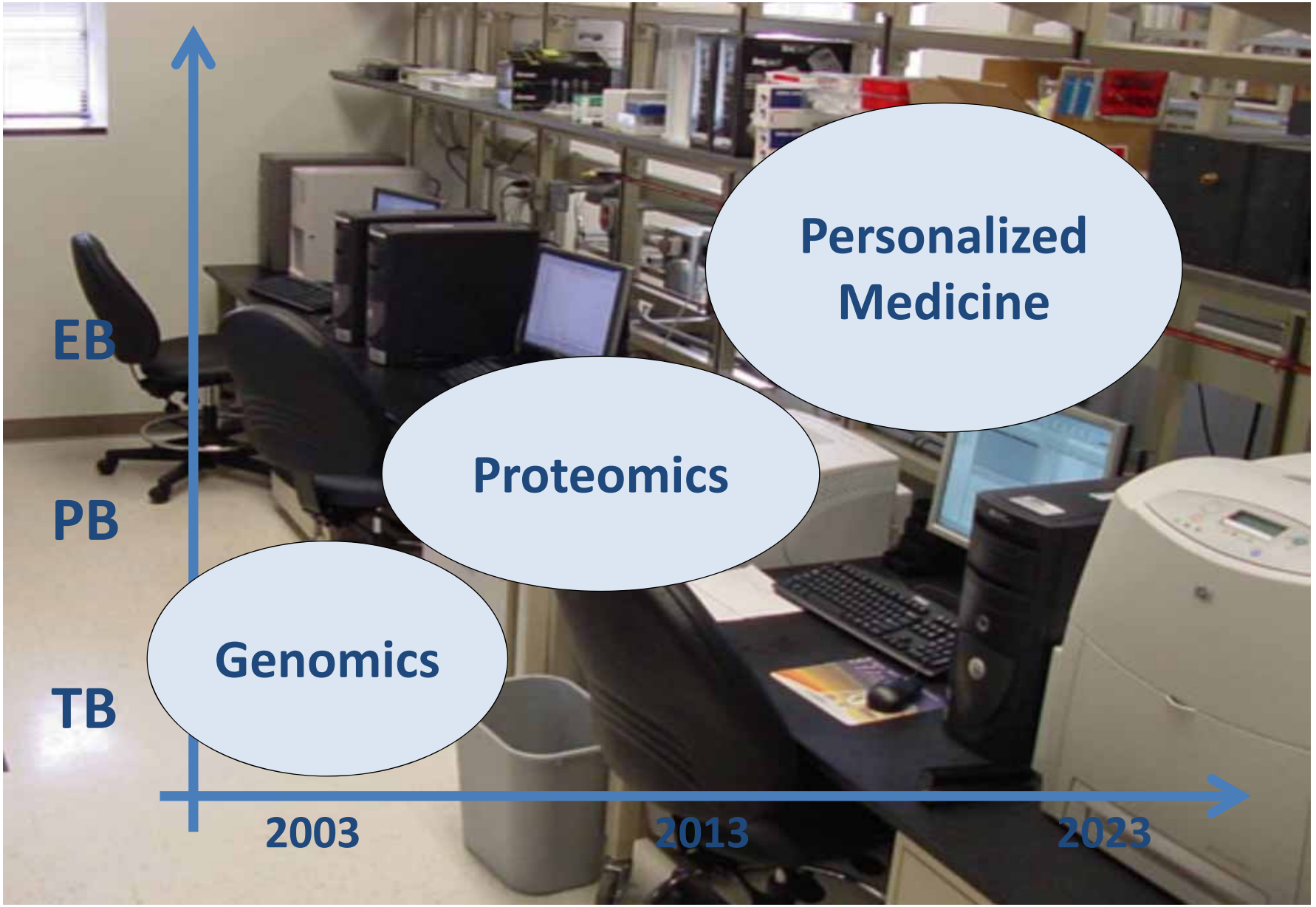
He, Y., Chen, Y., Alexander, P., Bryan, P. N. & Orban, J. (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. Proceedings of the National Academy of Sciences, 105, 38, 14412.




```
T0499  TTYKLILNLKQAKEEAIKEAVDAGTAEKYFKLIANAKTVEGVWTYKDEIKTFTVTE
       IIIIIIIIIIIIIIIIIIIIXIIIIIIIIIXIIIIIIIIIIIIIIIXIIIIIIIIIII
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.**

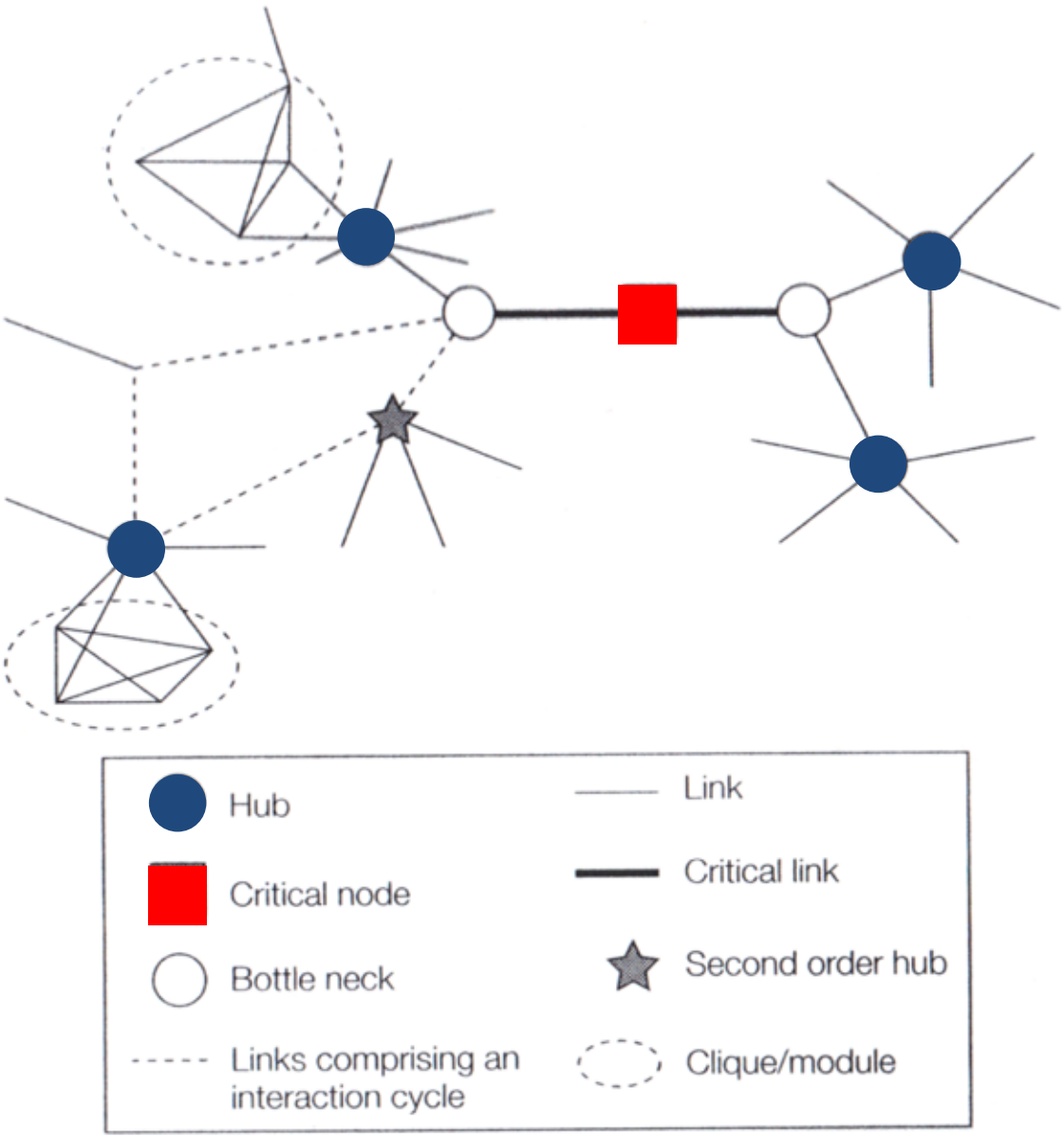

EB
PB
TB
Personalized Medicine
Proteomics
Genomics
2003
2013
2023

5) Review of 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

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

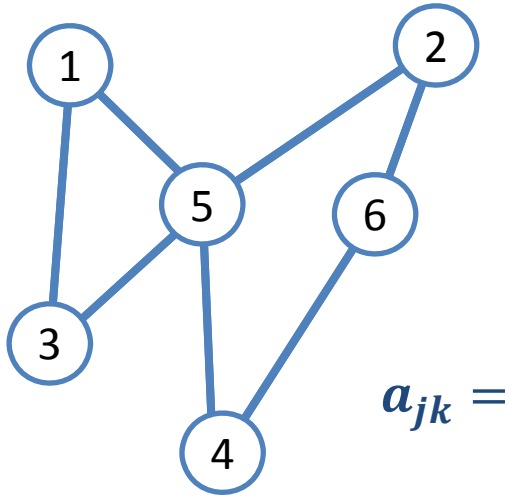


Hodgman, C. T., French, A. & Westhead, D. R. (2010) *Bioinformatics. Second Edition. New York, Taylor & Francis.*

Baby Stuff: Computational Graph Representation

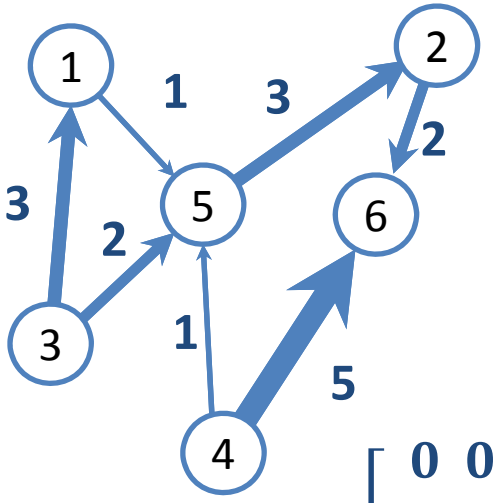
Adjacency (ə-ˈjā-sᵊn(t)-sē) **Matrix** $A = (a_{jk})$

$$a_{jk} = \begin{cases} 1, & if\ \{j,k\} \in E \\ 0, & otherwise \end{cases}$$



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

Simple graph, symmetric, binary

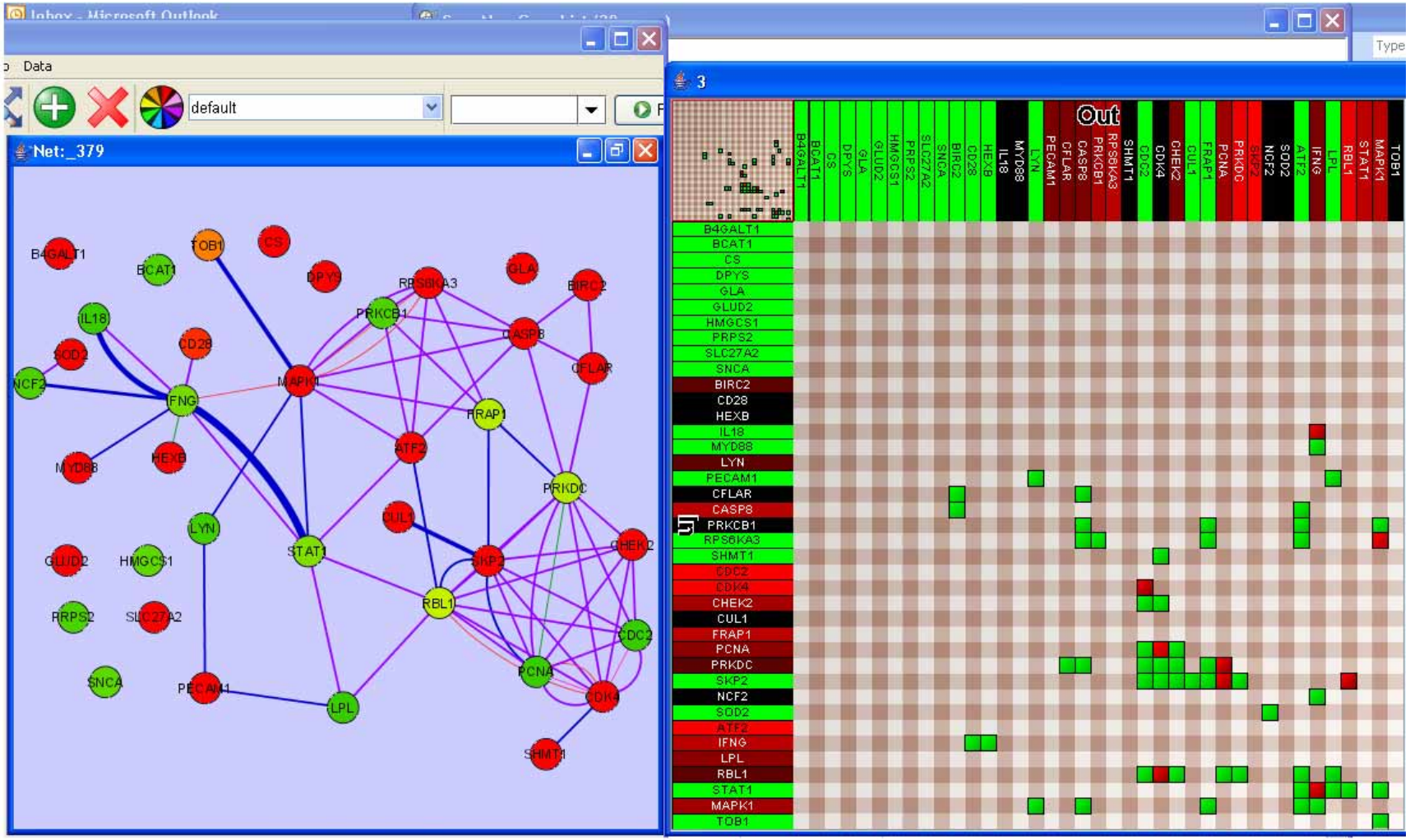


$$w_{jk} = \begin{bmatrix} 0 & 0 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \\ 3 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -5 \\ 1 & -2 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \end{bmatrix}$$

Directed and weighted

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

Example: Tool for Node-Link Visualization



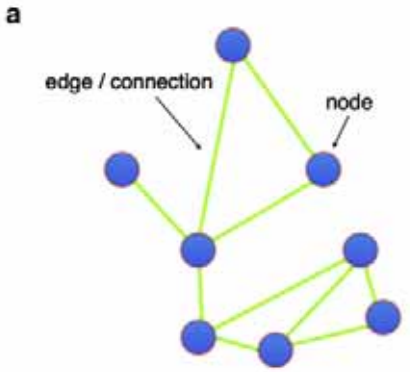
Jean-Daniel Fekete http://wiki.cytoscape.org/InfoVis_Toolkit

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

Some Network Metrics (1/2)

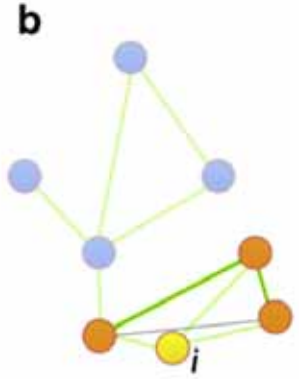
Order = total number of nodes n; **Size** = total number of links (a):

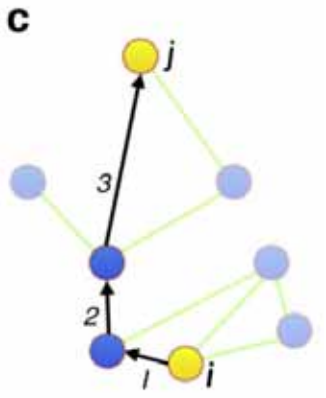
$$\sum_i \sum_j a_{ij}$$



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:

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





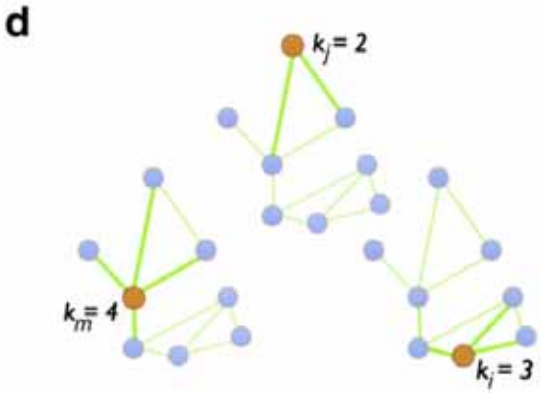
Path length (c) = is the arithmetical mean of all the distances:

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

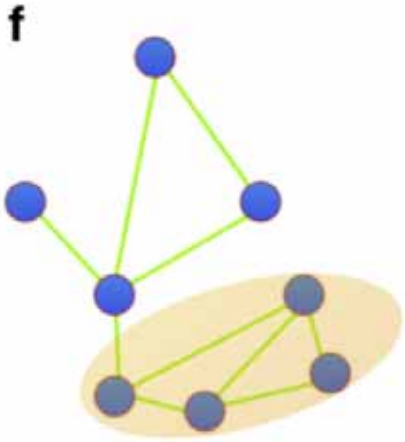
Costa, L. F., Rodrigues, F. A., Travieso, G. & Boas, P. R. V. (2007) Characterization of complex networks: A survey of measurements. *Advances in Physics, 56, 1, 167-242.*

Some Network Metrics (2/2)

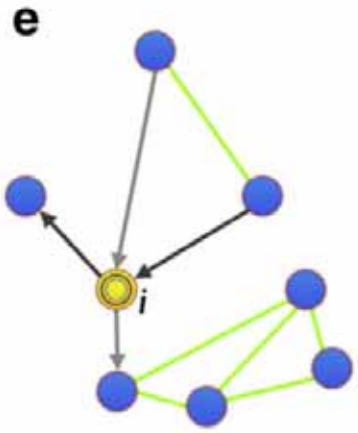
- **Centrality** (d) = the level of "betweenness- centrality" of a node I ("hub-node in Slide 28);



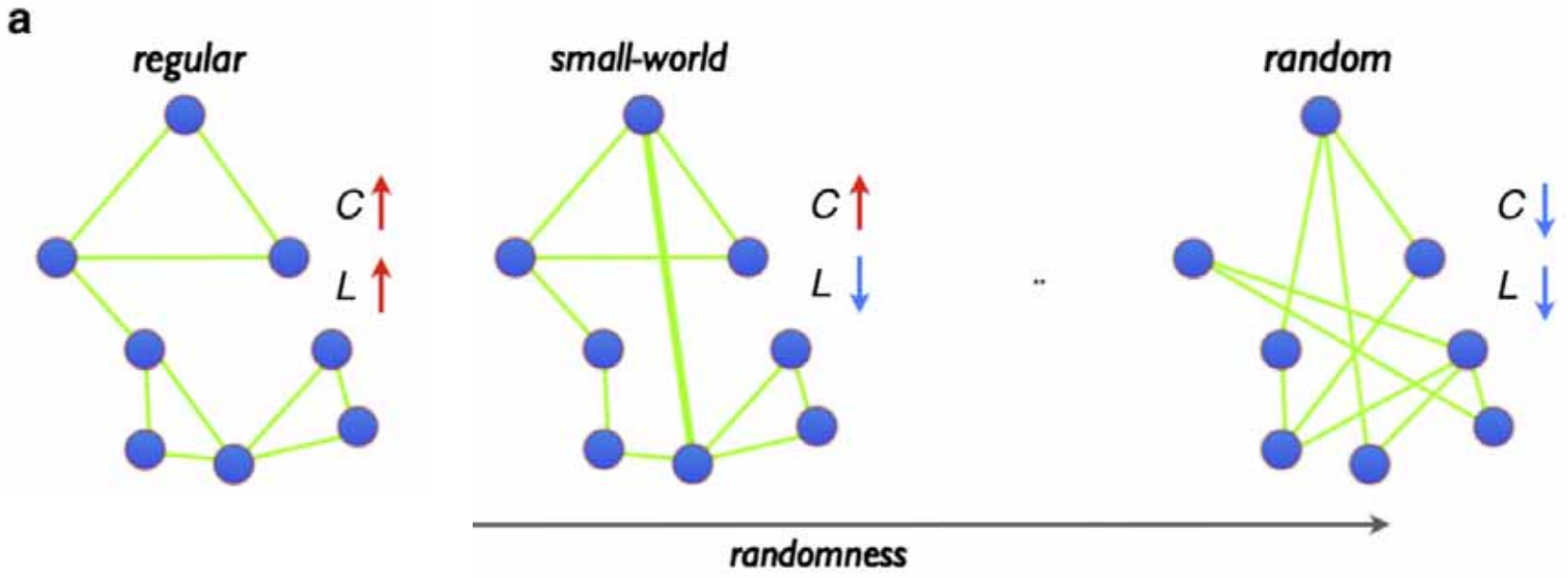
- **Nodal degree** (e) = number of links connecting *i* to its neighbors: $k_i = \sum_i a_{ij}$

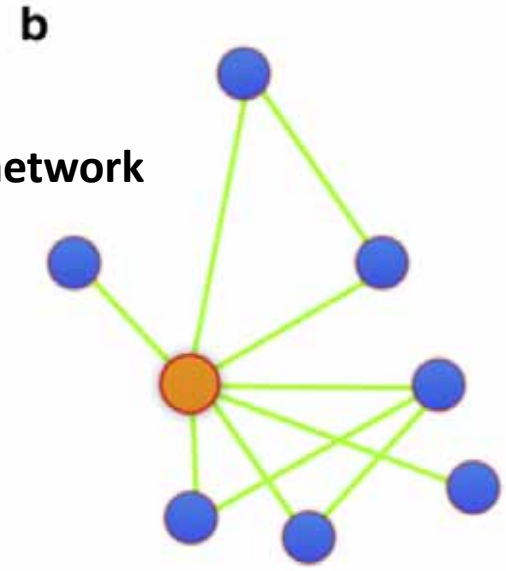


Modularity (f) = describes the possible formation of communities in the network, indicating how strong groups of nodes form relative isolated sub-networks within the full network (refer also to Slide 5-8).



Network Topologies

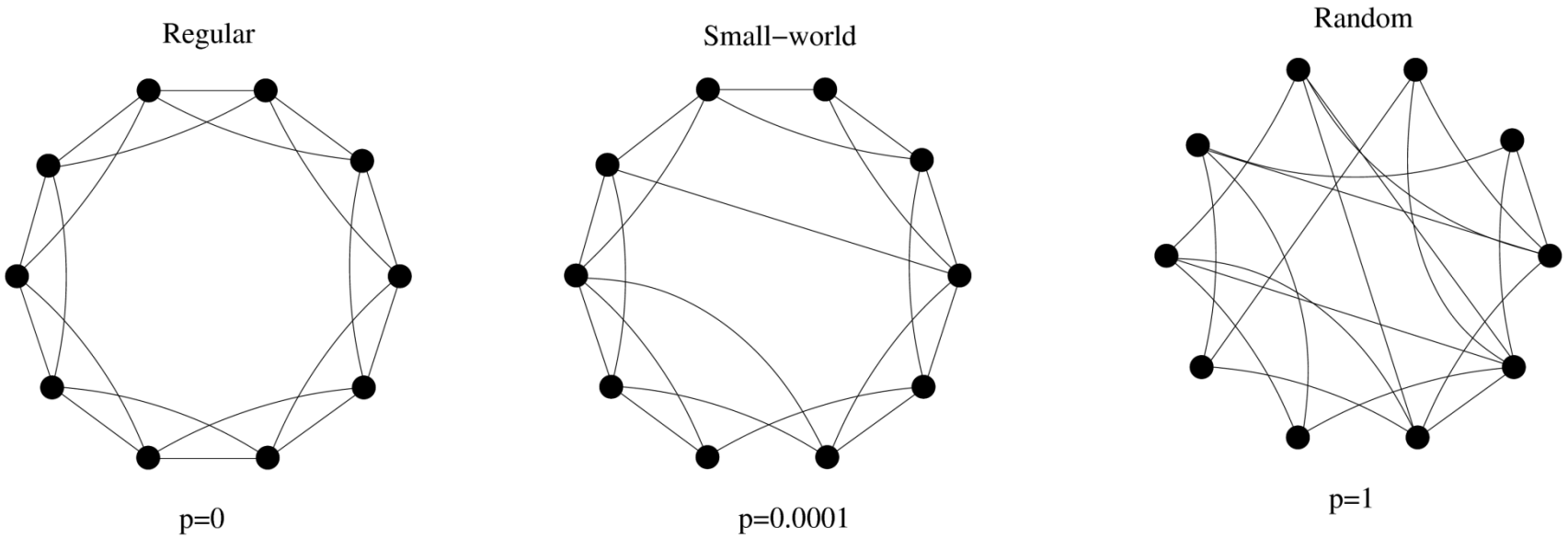




Scale-free network

Van Heuvel & Hulshoff (2010)

Small-World Networks

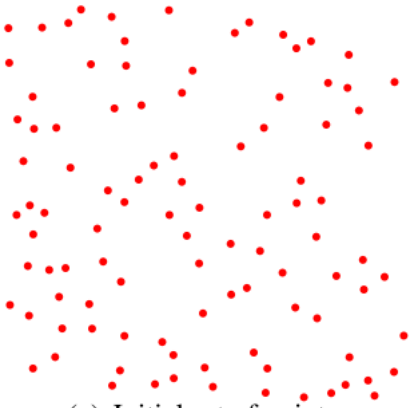


29.000 citations …

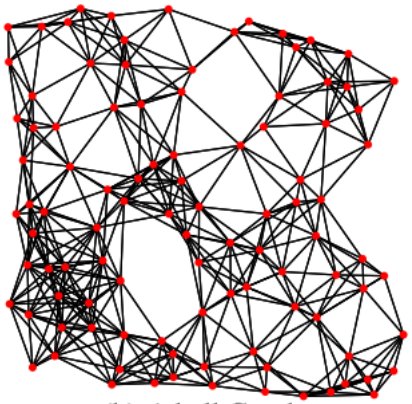
Watts, D. J. & Strogatz, S. (1998) Collective dynamics of small-world networks. *Nature, 393, 6684, 440-442.*

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

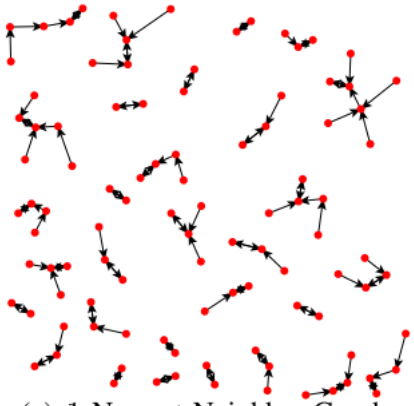
Slide 5-15 Graphs from Point Cloud Data Sets



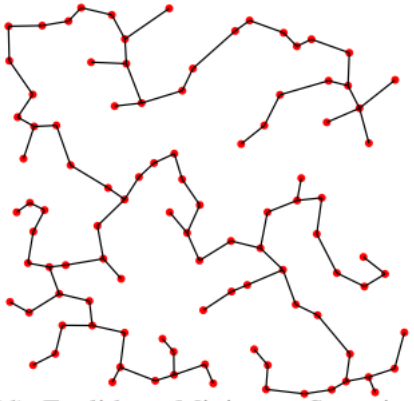
(a) Initial set of points.



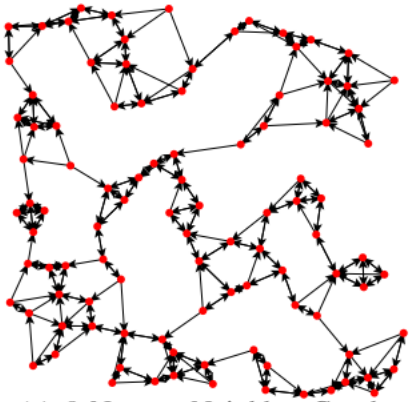
(b) 1-ball Graph.



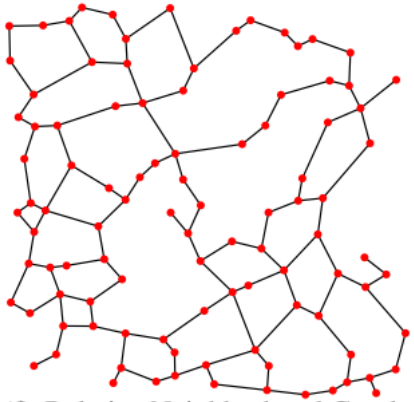
(c) 1-Nearest-Neighbor Graph.



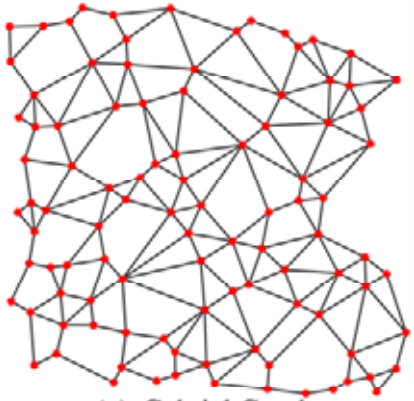
(d) Euclidean Minimum Spanning Tree.



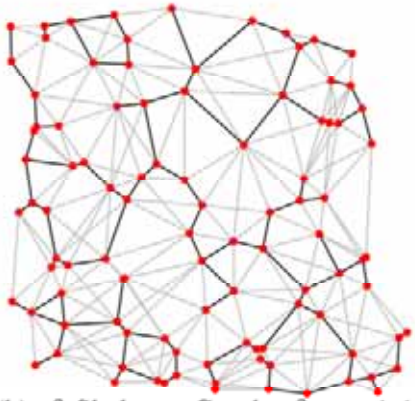
(e) 3-Nearest-Neighbor Graph.



(f) Relative Neighborhood Graph.



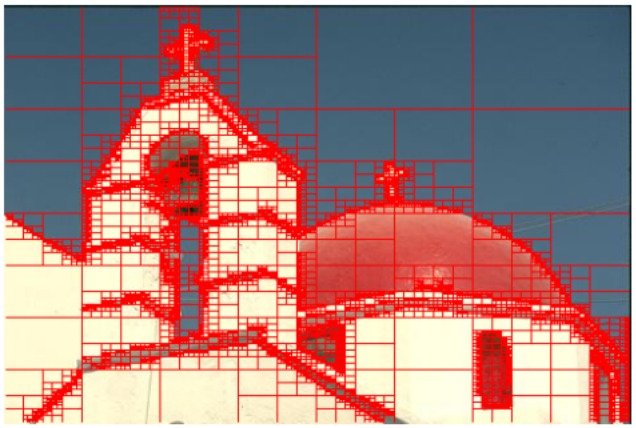
(g) Gabriel Graph.



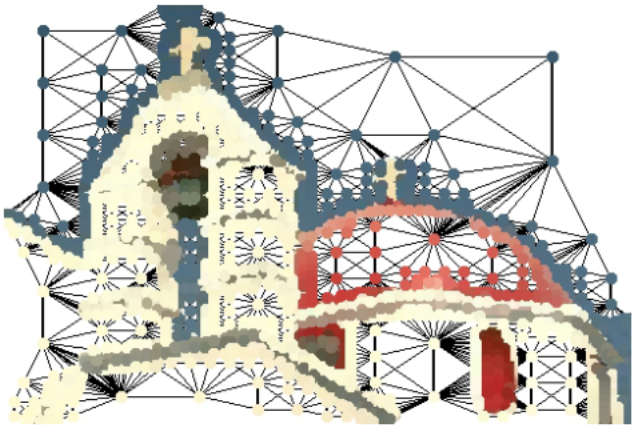
(h) β-Skeleton Graph, $\beta = 1.1$: black edges, $\beta = 0.9$: grey edges.

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.*

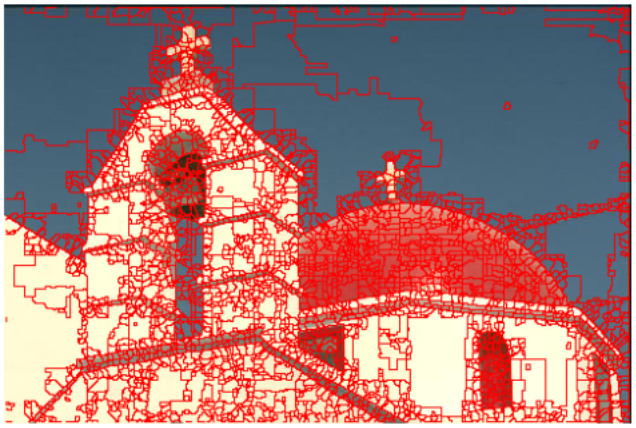
6) How do you get point cloud data from natural images?



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.*

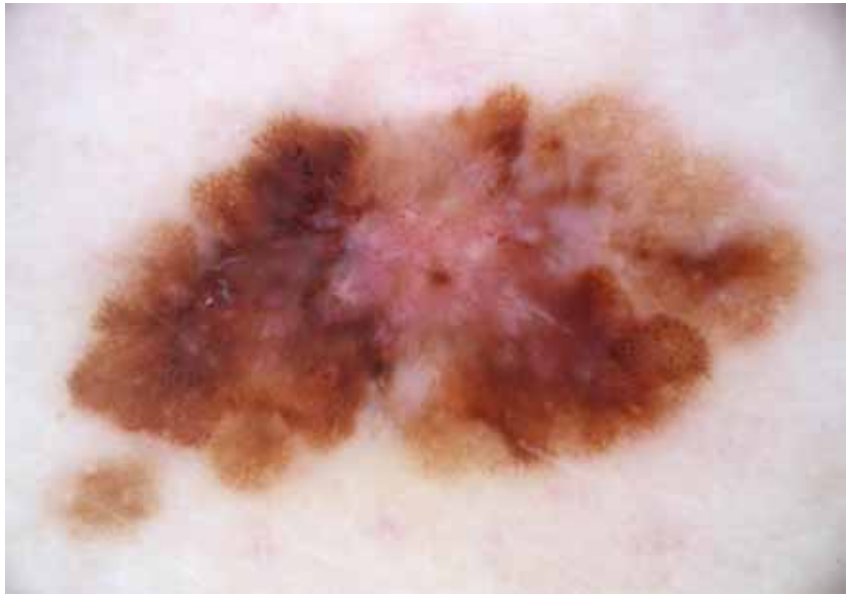
Example Watershed Algorithm

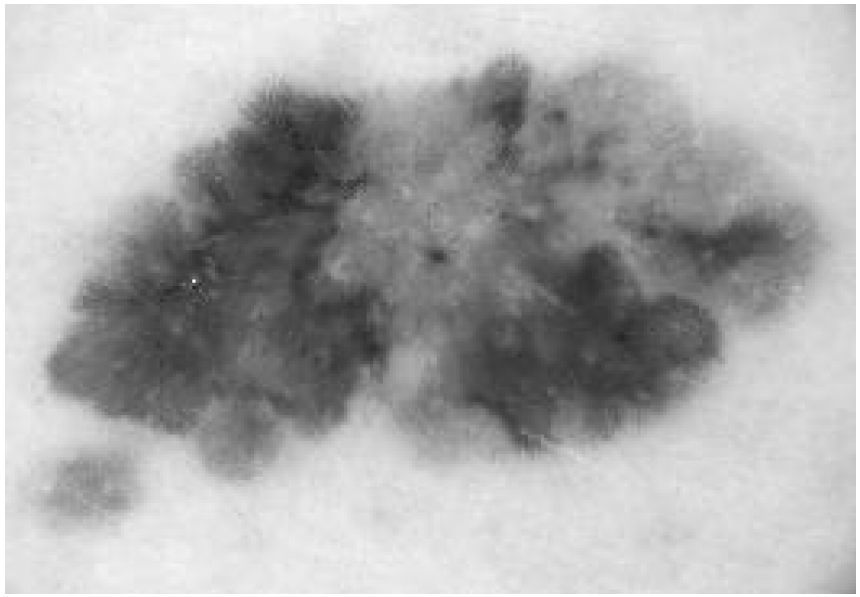
Algorithm 4.2 Watershed transform w.r.t. topographical distance based on image integration via the Dijkstra-Moore shortest paths algorithm.

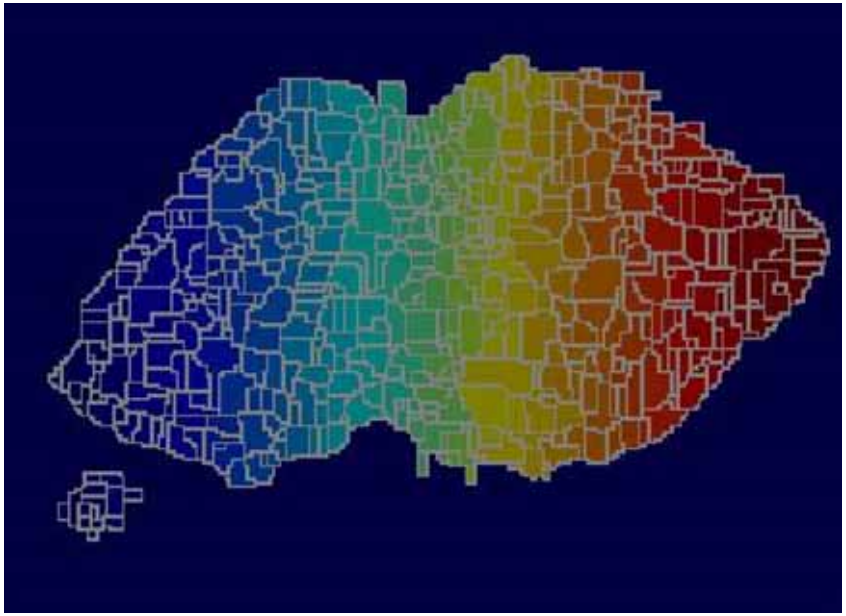
```
1: procedure ShortestPathWatershed;
2: INPUT: lower complete digital grey scale image G = (V, E, im) with cost function cost.
3: OUTPUT: labelled image lab on V.
4: #define WSHED 0                    (* label of the watershed pixels *)
5: (* Uses distance image dist. On output, dist[v] = im[v], for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

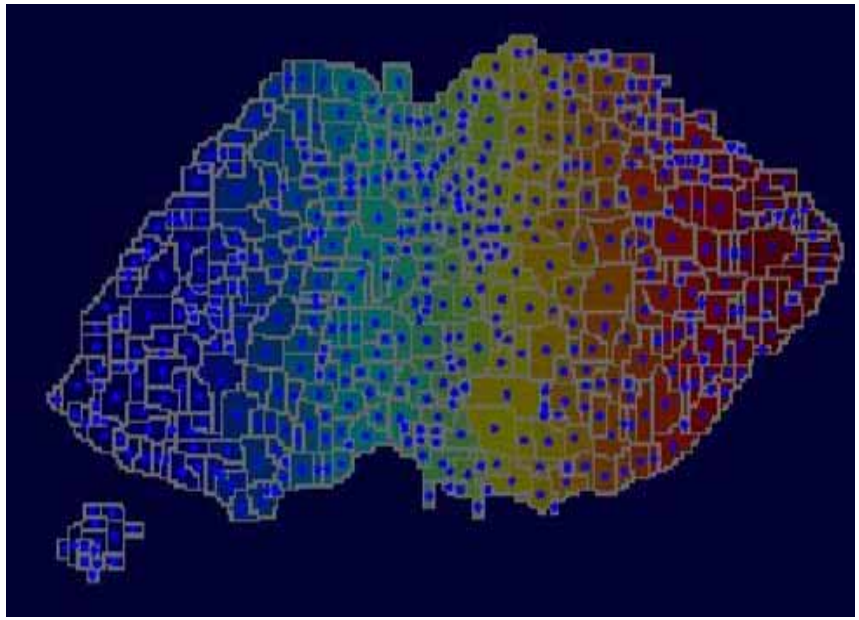
Meijster, A. & Roerdink, J. B. A proposal for the implementation of a parallel watershed algorithm. Computer Analysis of Images and Patterns, 1995. Springer, 790-795.

Graphs from Images: Watershed + Centroid

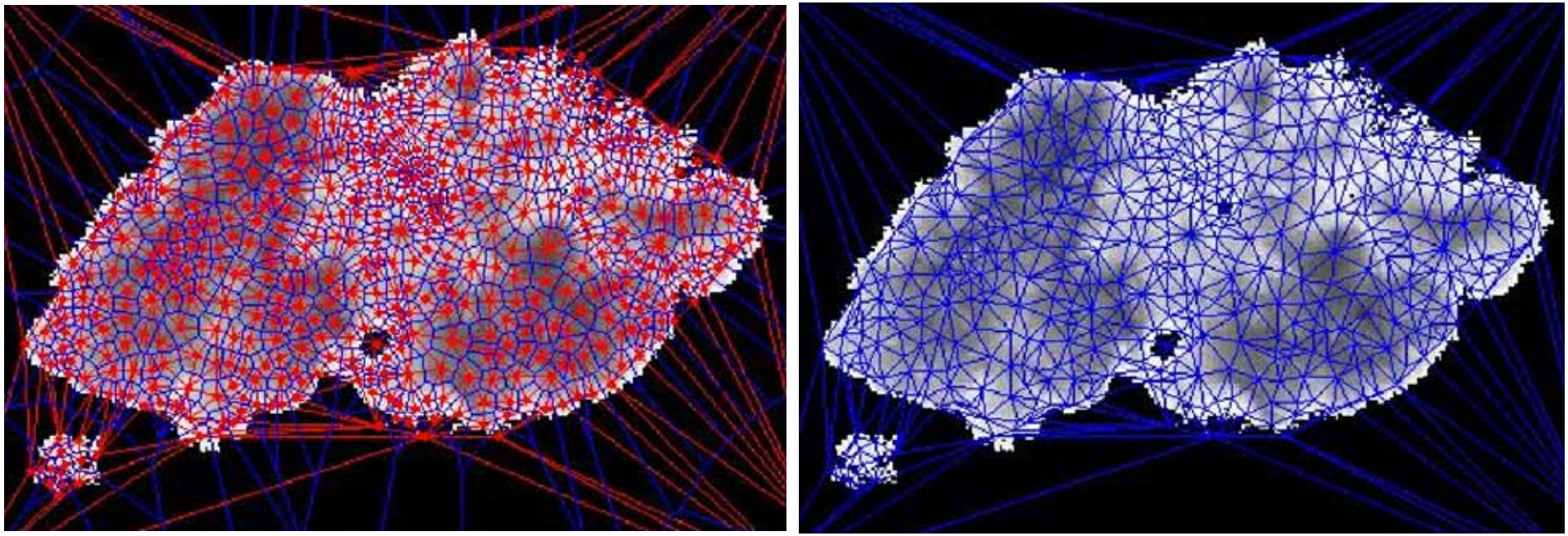








Slide 5-20 Graphs from Images: Voronoi <> Delaunay



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.

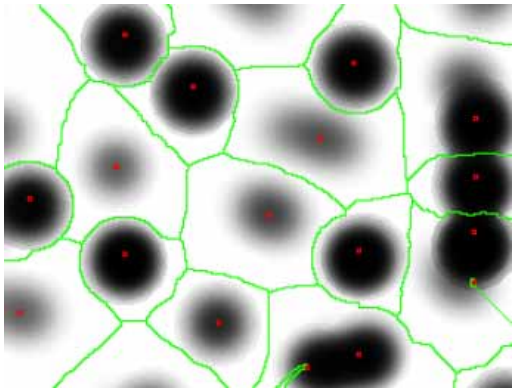
For Delaunay please refer to: Lee, D.-T. & Schachter, B. J. 1980. Two algorithms for constructing a Delaunay triangulation. Intl. Journal of Computer & Information Sciences, 9, (3), 219-242.

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

Watershed Algo based on connected components

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

Challenges

- 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**

State-of-the-Art Facts

- Study of complex networks started in the 1990s with the insight that real networks contain properties not present in random (Erdös-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).

7) Graphical Model Learning

- 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

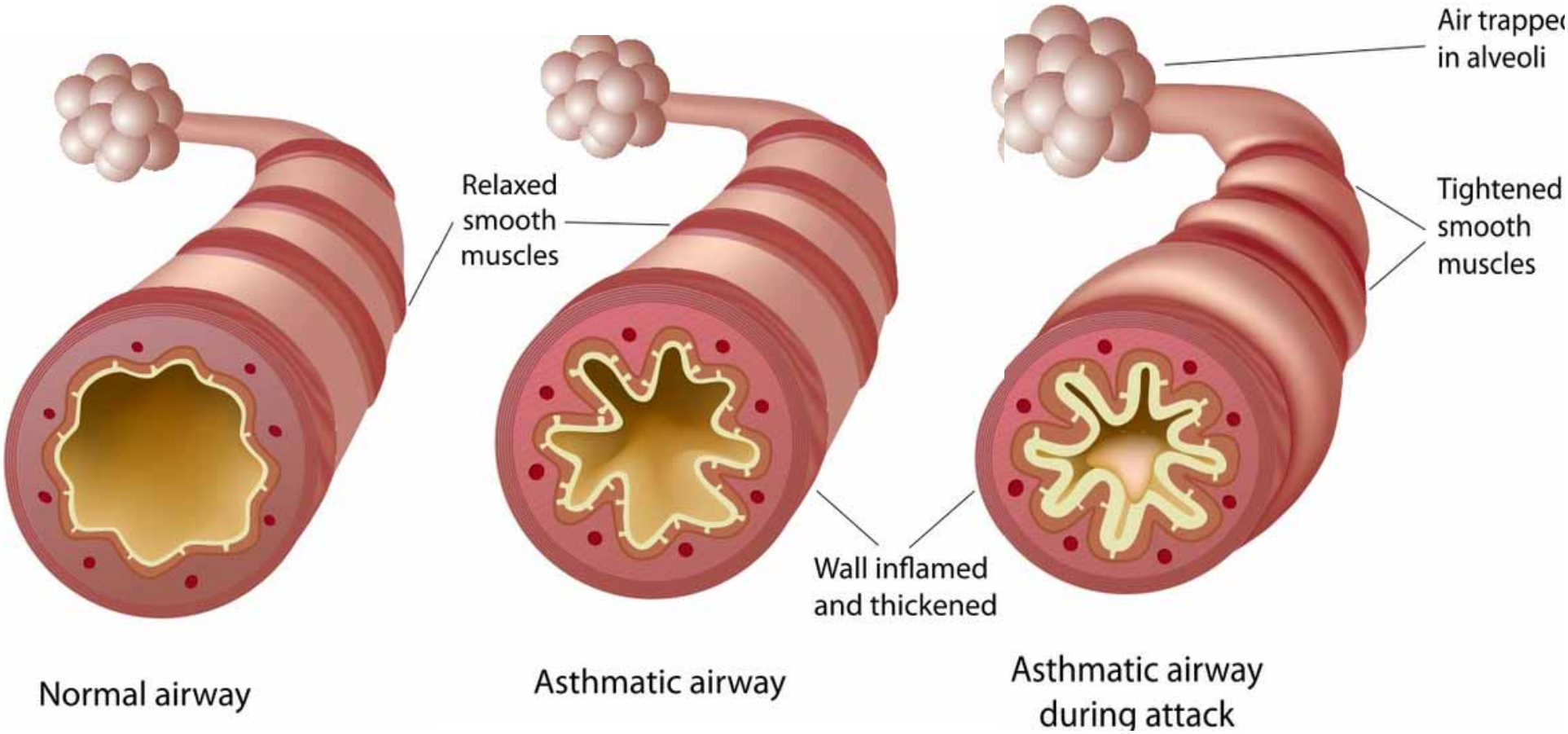
Jordan, M. I. 1998. Learning in graphical models, Springer

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?

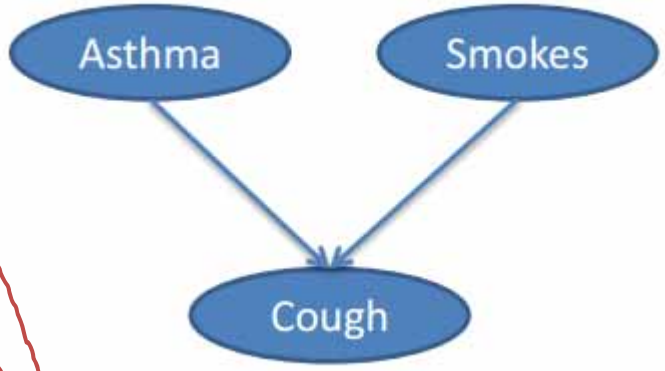


Beasley, R. 1998. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. The Lancet, 351, (9111), 1225-1232, doi:http://dx.doi.org/10.1016/S0140-6736(97)07302-9.

Example for Graphical Model Learning

Bayesian Network

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

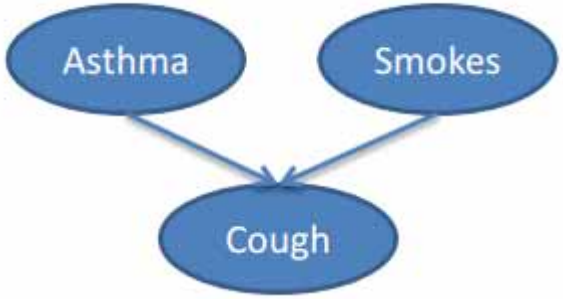


Florian	0	?	?

Florian	0	0.3	0.2

Rows are independent during learning and inference!

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



2.1 Asthma ⇒ Cough

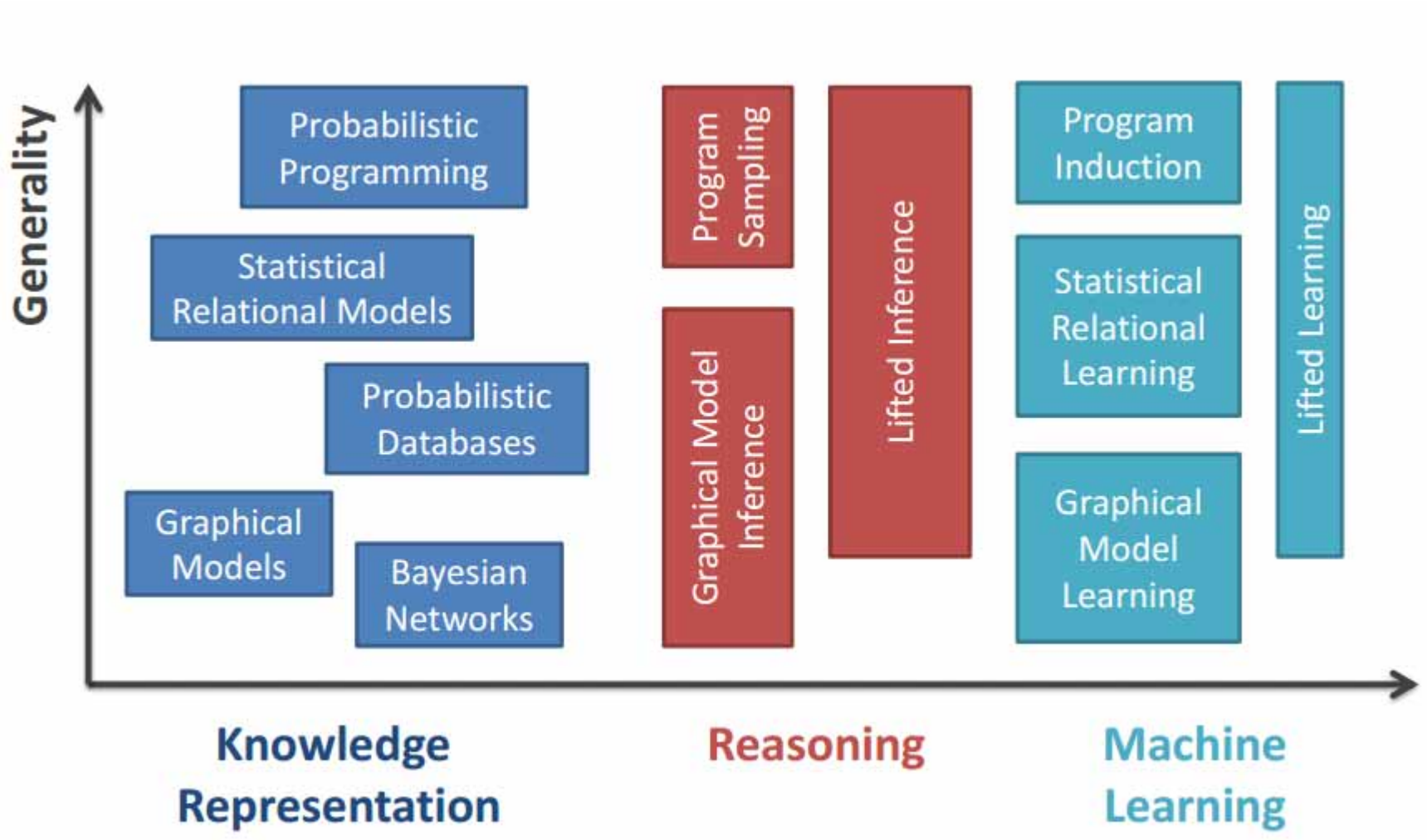
3.5 Smokes ⇒ Cough

2.1 Asthma(x) ⇒ Cough(x)

3.5 Smokes(x) ⇒ Cough(x)

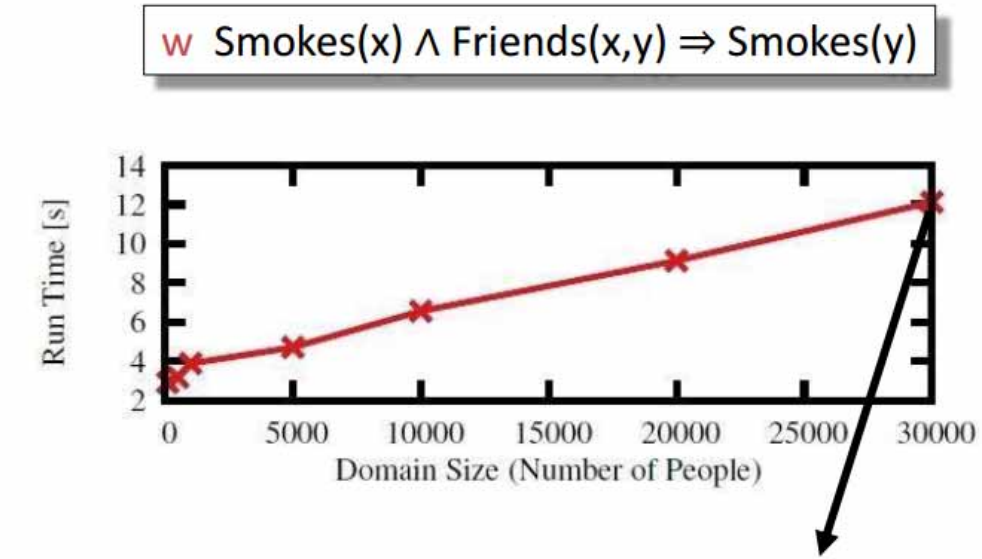
1.9 Smokes(x) ∧ Friends(x,y)
⇒ Smokes(y)

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



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).

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**



~~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.

Thank you!

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?