Neocybernetic Networks

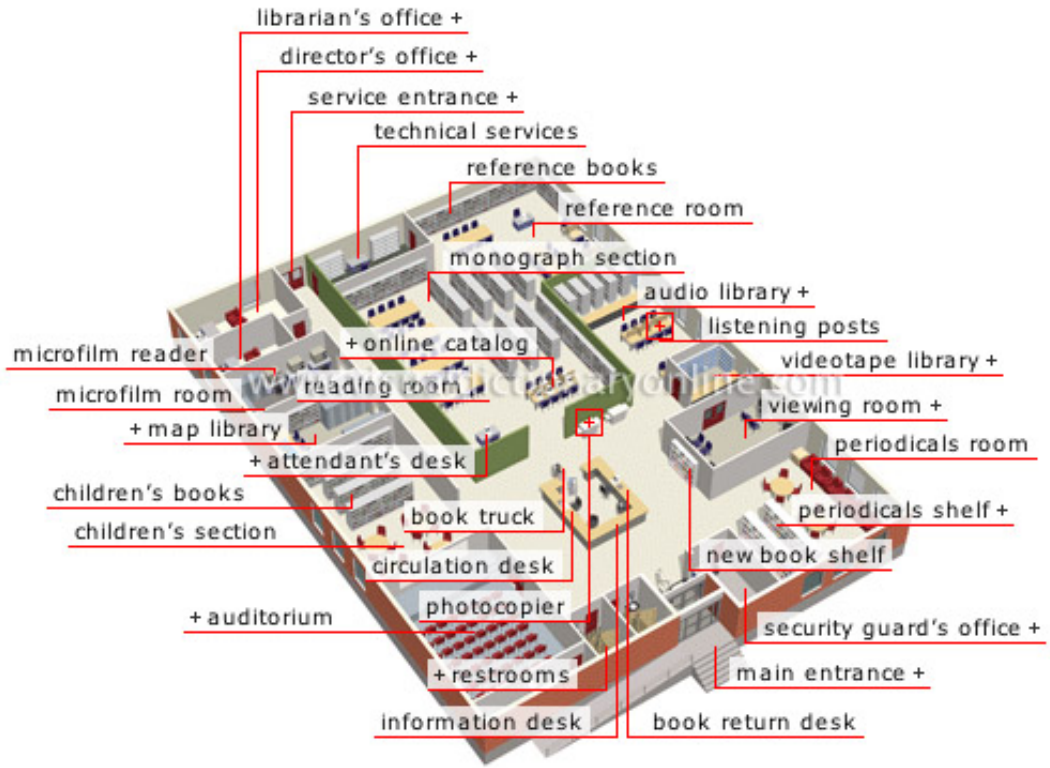
Presentation April 29, 2009
for course *Inf-0.3101* at TKK
by *Heikki Hyötyniemi*

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Automation and Systems Technology
Cybernetics Group

Example: development in *information search*

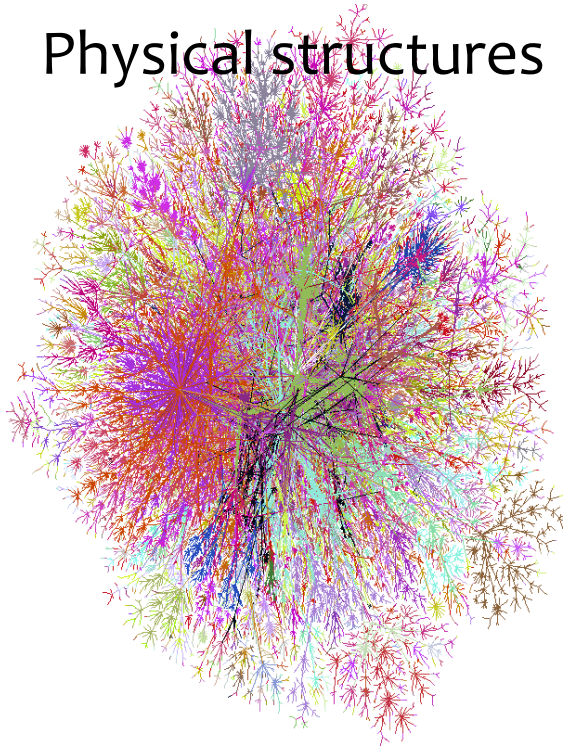
1. Fixed paths – Simply, *you just go to a library and do it*!
2. Logical paths – Hyperlinks in *Internet*
3. *Relevant* paths – Higher-level user interfaces like *Google* doing autonomous query-specific associations

- That means:
  From *mazes* towards information highways!

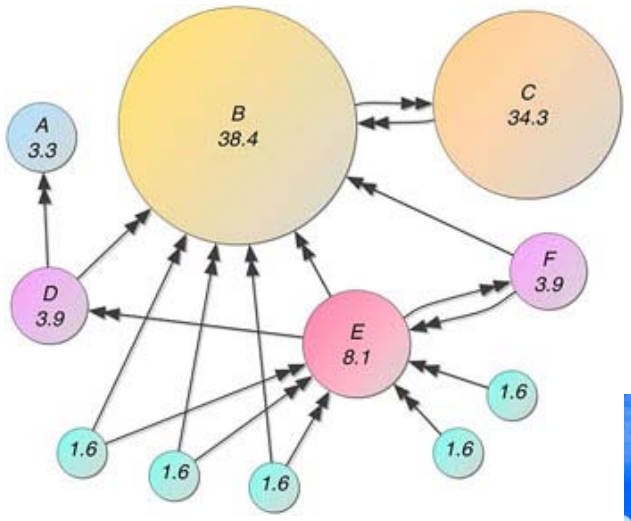


Towards higher-level views of all networks

Physical structures



Logical structures



"Functional structures"

- From information flows to "knowhowflows"
- From constraints to *freedoms*

Eternal (?) dilemma

- Familiar AI problem: How to reach automated intelligence?
- To implement *knowledge* one needs *semantics*
- *Contextual semantics* is delivered through couplings or *links*, and through other predetermined definitions
- But, for example, today's "Semantic Web" with its fixed ontologies is more like "sementic web" – there is *no learning* (what happened to those once celebrated *expert systems*?!)

- To reach the *functional* level, **pragmatism** needs to be applied as a guiding hand: How reality works in practice
- One should apply data mining in *inverse* direction – that is, learn from the best = *human behaviors*!

FUTURE!

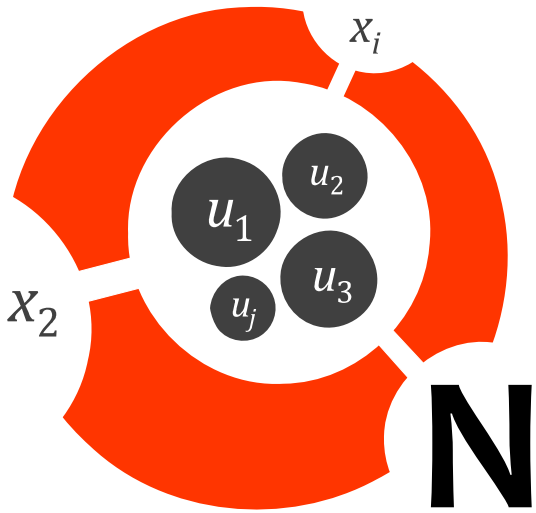
- What is more, the learning must be distributed – there cannot exist central control...

  = There must be **emergence**!

Neocybernetics: Approach to reach *emergence*

- Starting point: research on *neural networks*
- It has been observed that, given vectors of external excitations u and internal activivities x, local *Hebbian learning* in synapses can be characterized as $x = q\,\mathrm{E}\{xu^{\mathrm{T}}\}\,u$
- There is competition for activity as *exploitation* means *exhaustion*, implementing negative feedback through the environment, resulting in *self-regulation* and *self-organization*
- Self-organization: The neuron grid implements *principal component analysis* and *sparse coding* for input data
- If inputs are seen as resources, this strategy represents *evolutionary optimality* – extension to other fields, to general *distributed networked agent systems* becomes possible

- See **neocybernetics.com**



$$x_i \underset{\geq 0}{\overset{+}{=}} q_i \mathrm{E}\{x_i u_j\} u_j$$

Neocybernetics

$$\tfrac{1}{2} x^{\mathrm{T}} \langle x x^{\mathrm{T}} \rangle x - x^{\mathrm{T}} \langle x u^{\mathrm{T}} \rangle u$$

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Automation and Systems Technology
Cybernetics Group

- Based on solid mathematics...

  skipped here!

From (28) one can write yet another expression for the covariance:

$$\mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\} = \left(Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\right)^{-1} \mathrm{E}\left\{\bar{x}\bar{u}^{\mathrm{T}}\right\} \mathrm{E}\left\{uu^{\mathrm{T}}\right\} \mathrm{E}\left\{\bar{u}\bar{x}^{\mathrm{T}}\right\} \left(Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\right)^{-1}.$$

Eliminate the matrix inverses by multiplication, so that

$$\begin{aligned}\left(Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\right) \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\} \left(Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\right) \\ = \mathrm{E}\left\{\bar{x}\bar{u}^{\mathrm{T}}\right\} \mathrm{E}\left\{uu^{\mathrm{T}}\right\} \mathrm{E}\left\{\bar{u}\bar{x}^{\mathrm{T}}\right\},\end{aligned} \tag{40}$$

and observe the commutativity of the matrices:

$$\begin{aligned}&\left(Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\right)^2 \\ &= Q^{-1/2} Q^{1/2} \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}^{-1/2} \mathrm{E}\left\{\bar{x}\bar{u}^{\mathrm{T}}\right\} \mathrm{E}\left\{uu^{\mathrm{T}}\right\} \mathrm{E}\left\{\bar{u}\bar{x}^{\mathrm{T}}\right\} \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}^{-1/2} Q^{1/2} Q^{-1/2} \\ &= Q^{-1/2} \theta^{\mathrm{T}} \mathrm{E}\left\{uu^{\mathrm{T}}\right\} \theta Q^{-1/2},\end{aligned}$$

and, further, because of the diagonalizing properties of θ,

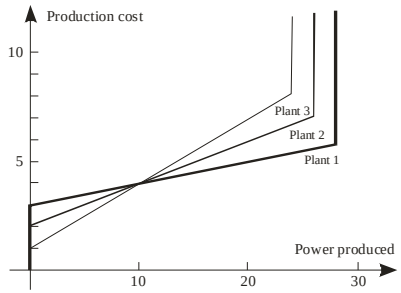
$$Q^{-1} + \mathrm{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\} = Q^{-1/4} \theta^{\mathrm{T}} \mathrm{E}\left\{uu^{\mathrm{T}}\right\}^{1/2} \theta Q^{-1/4}, \tag{41}$$

or

$$\mathrm{E}\left\{\bar{x}'\bar{x}'^{\mathrm{T}}\right\} + Q^{-1/2} = \theta^{\mathrm{T}} \mathrm{E}\left\{uu^{\mathrm{T}}\right\}^{1/2} \theta, \tag{42}$$

Design of networks: Case *energy production*

- Strict optimality:



- Predetermined profiles:

$$J' = \left(u - \varphi x\right)^T \left(u - \varphi x\right)$$

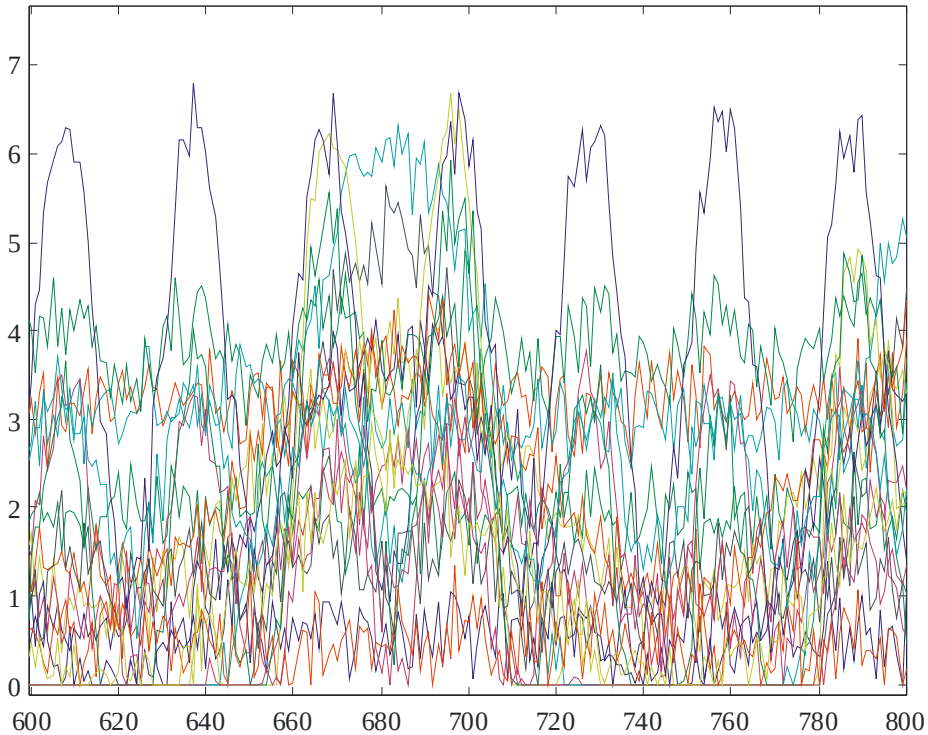
- Cybernetic cost:

$$J'' = \left(u - \phi x\right)^T \mathrm{E}\left\{uu^T\right\}\left(u - \phi x\right)$$

- Additional constraint:

$$\sum_{i=1}^{3} \overline{x}_i = \sum_{j=1}^{20} u_j$$

Behaviors of 20 consumers (u_j)



Goal: Optimize production of the three production units!

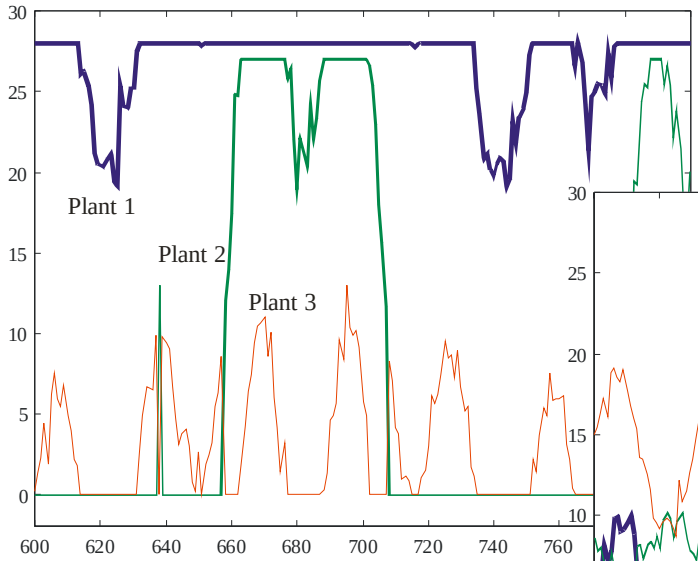
HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Automation and Systems Technology
Cybernetics Group

- Static minimization of the criterion separately for each time instant – three strategies experimented:

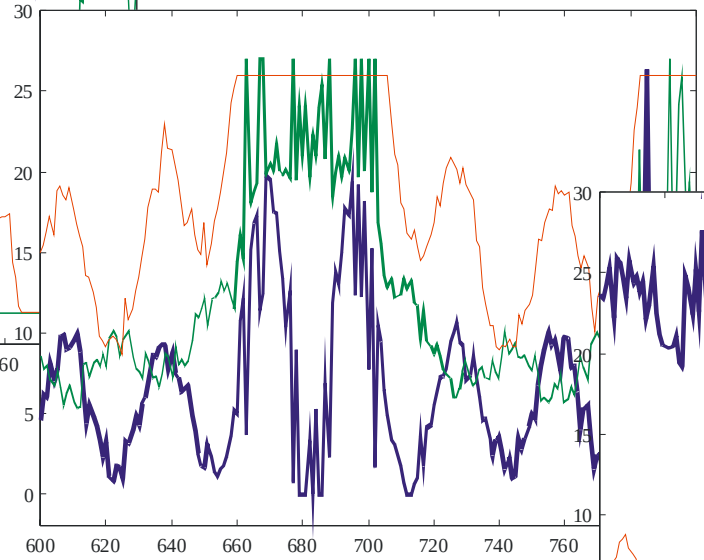
- Explicit optimization: Piecewise linear cost criterion means that only one of the producers is active at a time, others being in either of the extreme values (zero or maximum)
- Explicit distribution: Profiles φ define (randomly) preferred consumers for each producer; further, some plants can be "spare plants" to substitute malfunctioning master plants
- Cybernetic strategy: Profiles ϕ are determined by the correlation structures among consumers; because of the nonlinearities, there exist various minima to choose from

Quadratric optimum is rather flat – solutions in the vicinity of optimum are not much worse

Behaviors of 3 producers $(\bar{x}_i)$

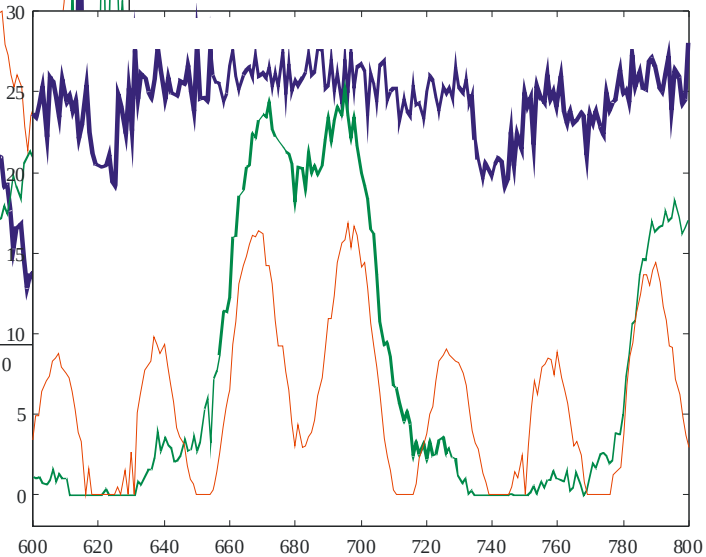


"Optimized"



Distributed

Cybernetic local balance near the explicit optimum



Cybernetic

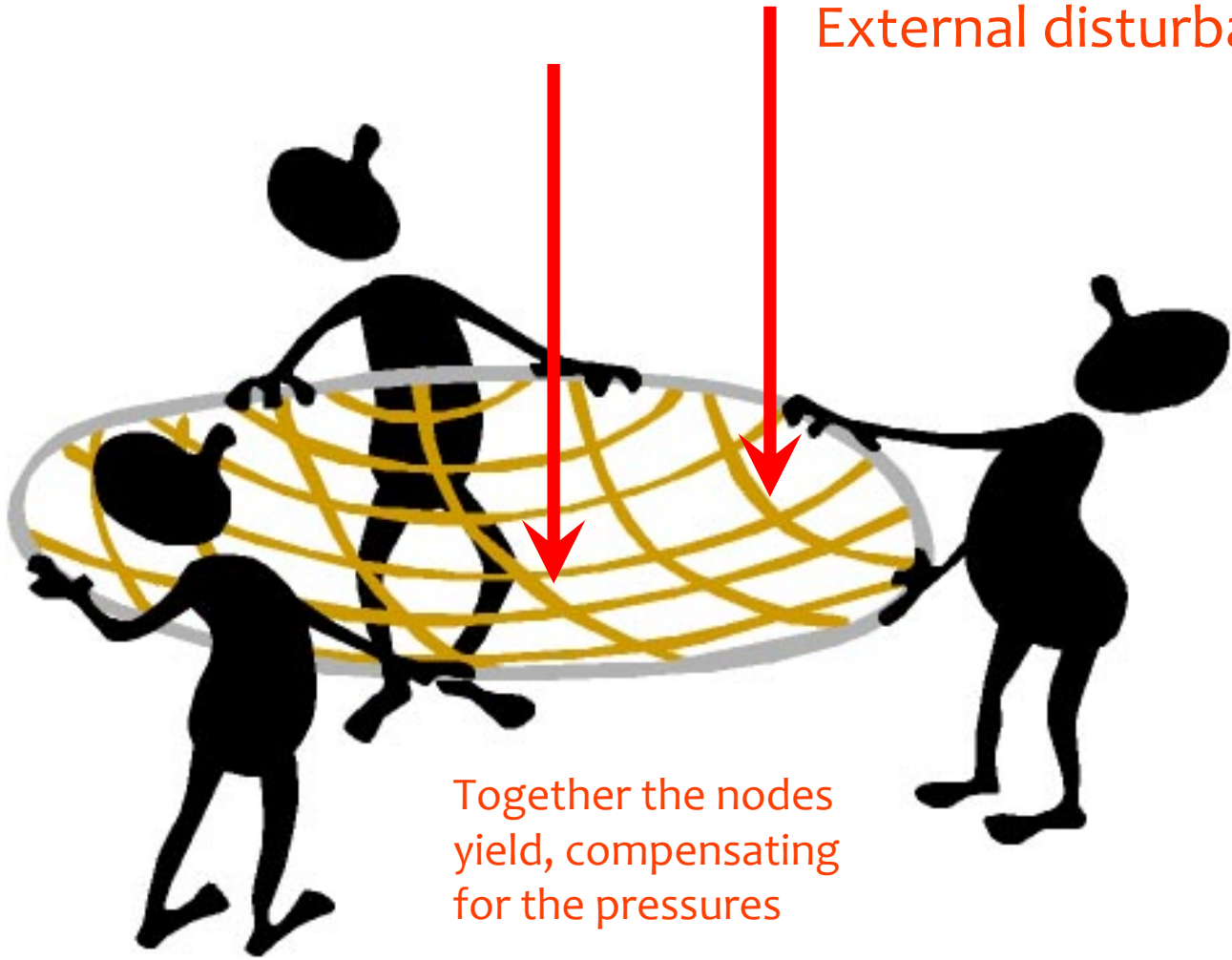
- There is a plenty of variation in the cybernetic case, but the variations are small = **robust?**

HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Automation and Systems Technology
Cybernetics Group

Technical networks in general

- Typically, the nodes in practical networks are not identical – they can have different roles, and these roles have to be taken into account in modeling
- The networks themselves are also very different:
  - In Internet, the "raw material" can be produced and copied indefinitely, restrictions and costs coming from transfer capacity
  - In power production, on the other hand, energy transfer is no problem, capacity restrictions and costs being caused in production
  - Still, the same modeling approaches can be applicable in both cases applying the idea of *dual graphs*?
- Possible applications: steam (pressure) pipelines in paper mills; design of electric networks with varying loads, etc.

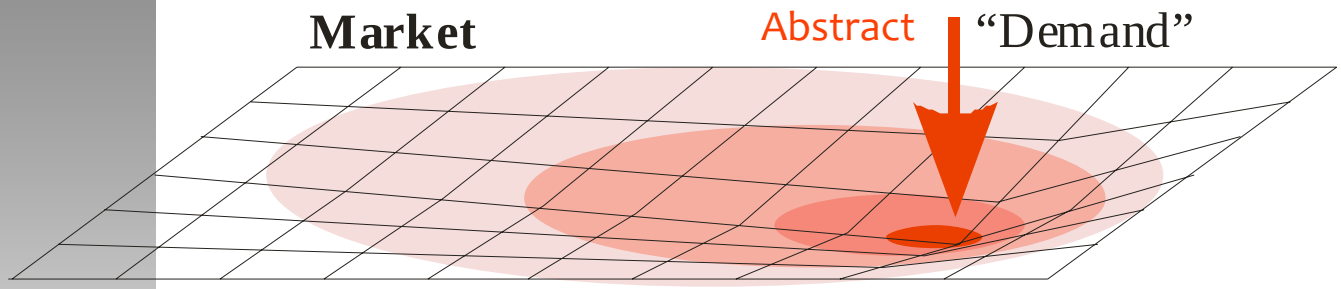
Cybernetic network behaves like a “safety net”



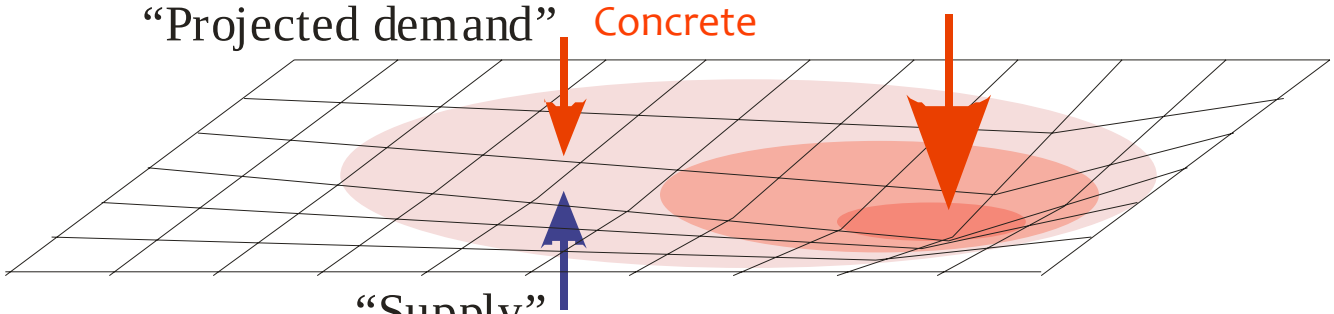
Analogy example: Supply and demand

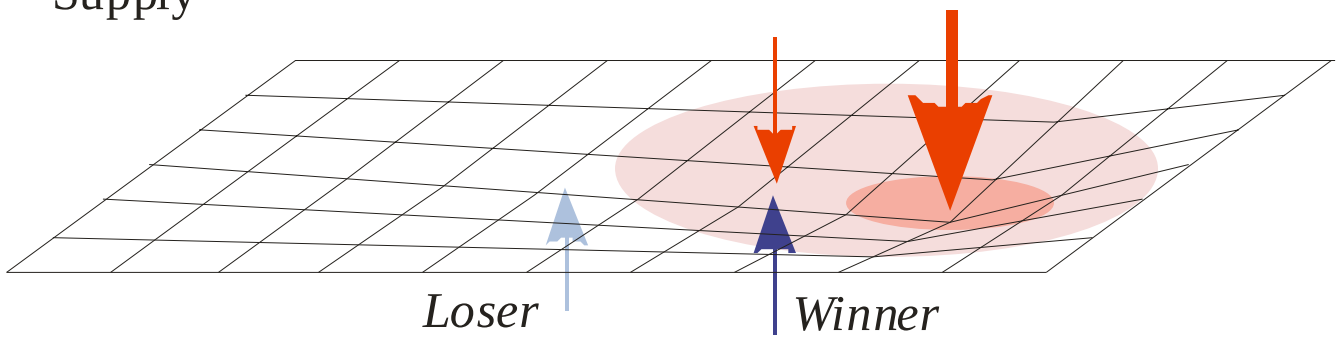
- Market (the system) is poorly structured and unknown; customer "needs" (external forces) and their realizations are not known, product substitution is not known
- Properties of products determine their "location" within the (infinite dimensional) market structure; there are many competitors with differing product properties
- The products offer the mechanism for compensating the external pressures: Deformation stands for the demand for the product (in terms of money available), x_i stands for supply (in terms of investments)?
- Balance always found; maximum overall benefit reached when applying cybernetic strategy?!

Market as a "rubber membrane"



Compare to *decoherence* – collapse space of possibilities





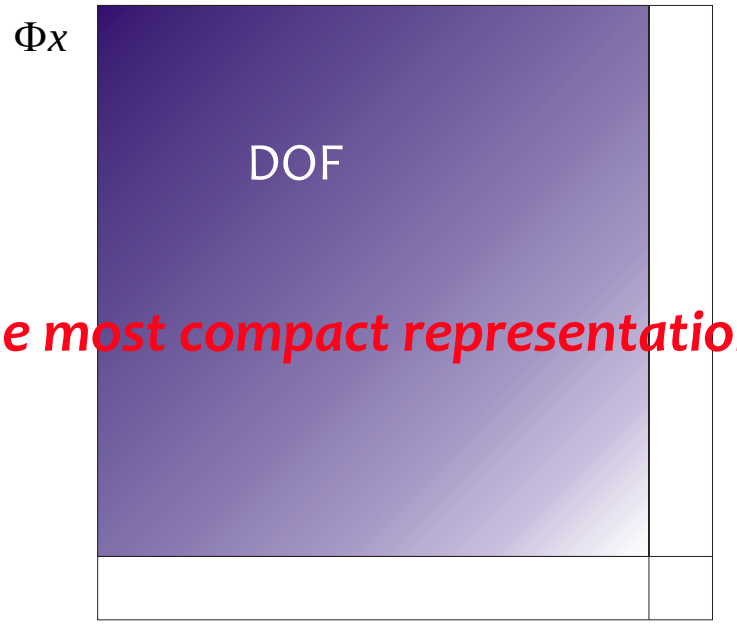
Constraints vs. *freedoms*

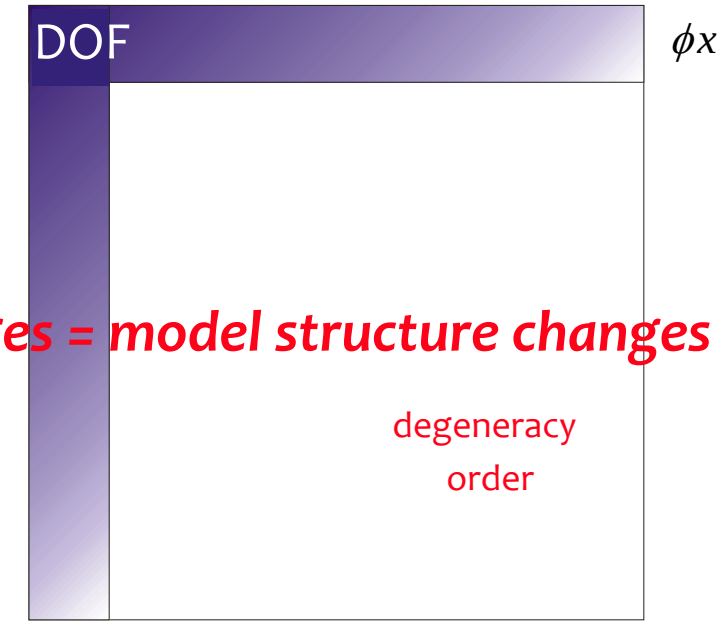
- The overall network structure determines in which directions there can be movement and where not
- Claim: The degrees of freedom are more characteristic to a system than the constraints are
- The constraint model determines a line in the data space – "null space", where there is no freedom among data
- "Axes of freedom" = remaining subspace that is *orthogonal to the null space* = basis of a NEW MODEL STRUCTURE
- The eigenvalue decomposition of the data covariance matrix reveals in which directions there is variation in the data and how much: *Eigenvectors = axes of freedom*, and *eigenvalues = their relevances*

View from above: "Emergent Models"

- Data high-dimensional
- Few connections = constraints
- Many degrees of freedom left

- Data equally high-dimensional
- Many constraints
- Few degrees of freedom (right!)





The most compact representation changes = model structure changes

For example, $a \leftrightarrow b$

$c \leftrightarrow d \quad d \leftrightarrow a$
$b \leftrightarrow c \quad c \leftrightarrow b \quad a \leftrightarrow d$
$a \leftrightarrow b \quad b \leftrightarrow d \quad d \leftrightarrow e$
$e \leftrightarrow c$
$a \leftrightarrow c \quad \ldots$

Approaches to networks

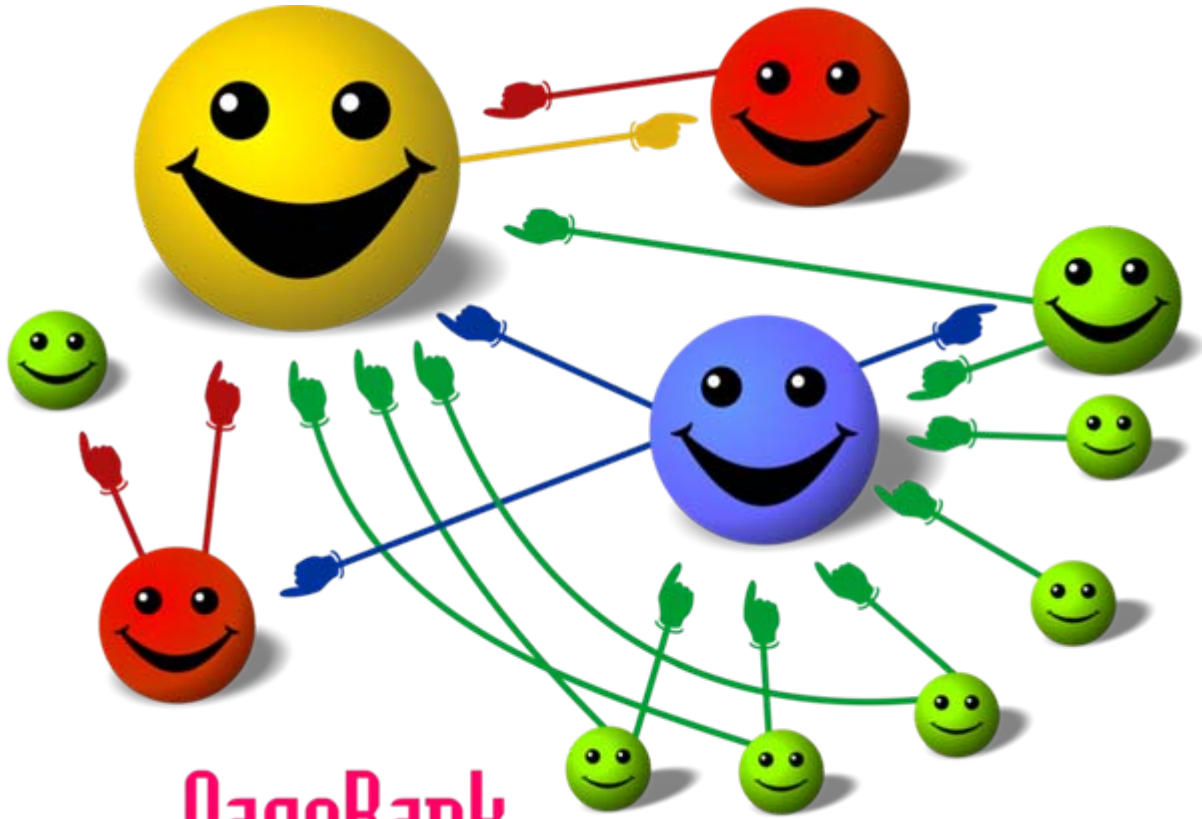
- Graph theory
  - Connections between nodes are “crisp”
  - However, there is a continuum of interaction effects: The connections in reality are not of “all-or-nothing” type
- Bayesian networks
  - Strong probabilistic theory – *assuming that assumptions hold…*
  - However, the “nodes” in real networks are often not independent of each other: Loops and alternative paths exist in complex networks
- Now: **Neocybernetic framework**
  - Numeric, non-crisp connections, **fully connected**
  - **“Pancausality”** taken as the starting point: It is assumed that, in equilibrium, **all** nodes are causes and **all** are effects – opposite approach!

Contribution of neocybernetic views?

- Abstract over individuals spatially and temporally
- Cybernetic system is a complex system that is characterized by dynamic equilibrium among opposing tensions
- The balances characterize dynamic attractors that are visible in the data and thus relevant in that domain
- Interacting systems are reactive, controlling each other, the overall dependencies becoming pancausal
- The system gets towards better and better coupling with its environment, meaning more fluent information flow
- During evolution (natural or not) the controls become more and more stringent and the overall system becomes stiffer
- Final result: "Degrees of freedom get eliminated" – **WHAT?**

Case: The Secret of Google

- The page ranking values are the entries of the *dominant* ***eigenvector*** of the (modified) adjacency matrix…



HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Automation and Systems Technology
Cybernetics Group

See e.g. Kurt Bryan, Tanya Leise (2006): The $25,000,000,000 Eigenvector: The Linear Algebra behind Google. *SIAM Review*, *48*(3): 569–581.

Neocybernetic interpretation

- Assumption: the search system becomes more and more optimal = more and more cybernetic!?
- Everything depends on what is seen as important and how variables are selected = "network semiosis":
  - Individual searches = "resources" u that nodes compete for
  - Visits, search paths = "activities" x of page providers
- Local actors try to maximize search depths, simultaneously exhausting the query, meaning that there is competition...
- This means that Hebbian learning finally optimizes the system structure – but this can be done explicitly, too?!
  - Neocybernetic equalization: the average "deformation" or "variance" in each direction is the same = typical searches u have shorter paths x
  - And, as the coupling tightens, the average search paths get shorter

Conclusion?

- Freedoms define the directions where variations “make a difference that makes a difference” (G. Bateson)
- Traditionally: constraints – world as it is / has to be
- Cybernetically: freedoms – “world as it could be”
- Neocybernetics gives new intuitions of innovation, where to do “nextworks”!