Lecture 11

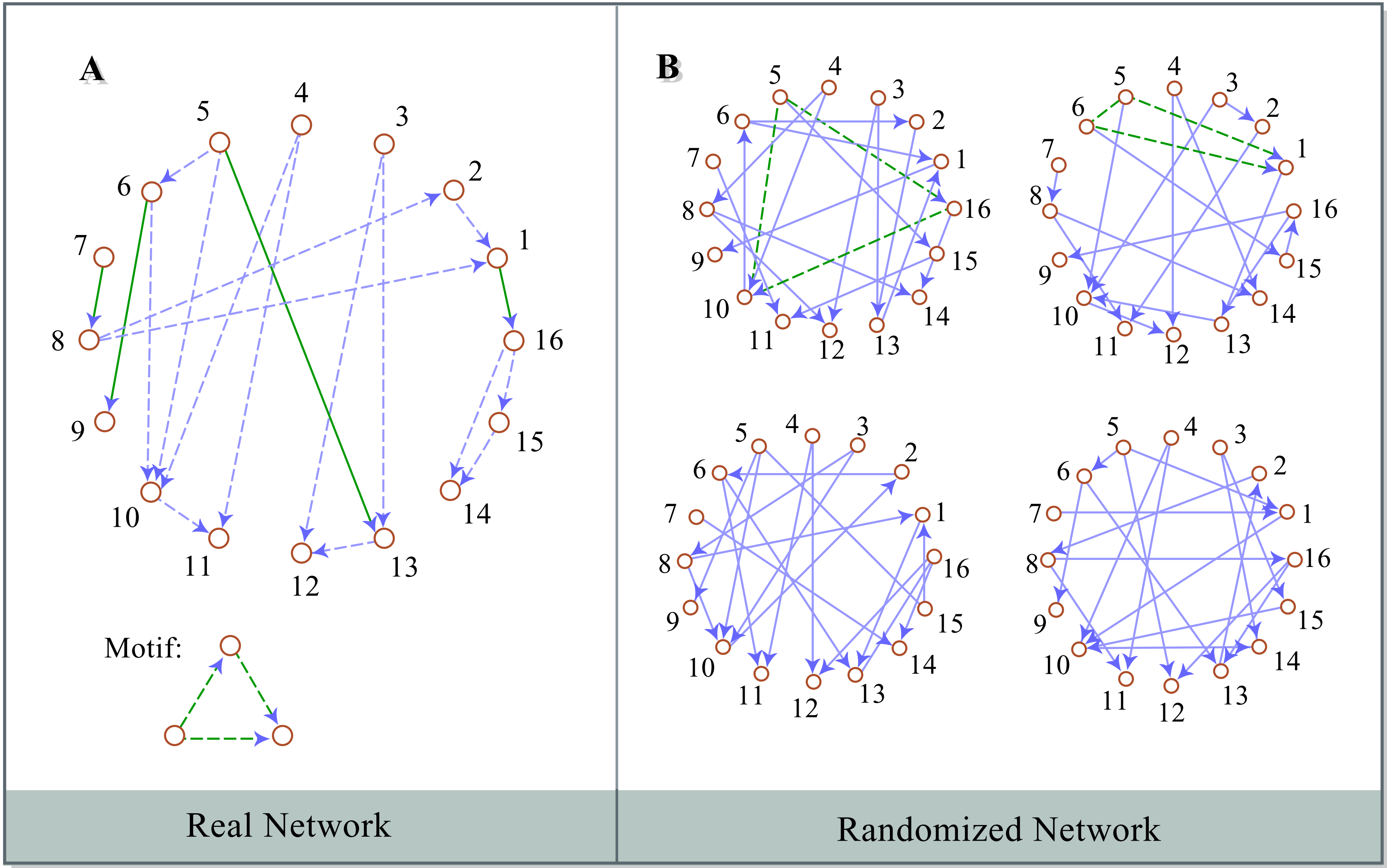
- Network terminology and structural characteristics
- Motifs (patterns of directed -and non-directed- links and a connection to **function**)
- A Complex system representation: **hierarchy of function**
- Coarse-Graining (abstractions of **function hierarchically** described) and PGNM
- Return to **modularity** discussion
- (Introduction to models (lecture 12 material as time permits)
- Electric Power student team report #1

Network Analysis Terminology -notated

- node (vertex), link (edge) CM2,
- size, sparseness, metrics CM2
- degree, average degree, degree sequence DW4
- directed, simple DW4
- geodesic, path length, graph diameter DW4
- **transitivity (clustering),** DW10 connectivity, reciprocity CM6
- centrality (degree, closeness, betweenness, information, eigenvector) CM6
- prestige, acquaintance CM6
- ideal graphs (star, line, circle, team) CM6
- degree distribution, power laws, exponents, truncation, CM6
- Models (random, "small world", poisson, preferential attachment)
- constraints, rewiring DW7
- Hierarchy DW7, JM8&9, CM11
- **community structure**, cliques, homophily, assortative mixing, **degree correlation coefficient** DW10
- **motifs, coarse- graining** CM11
- navigation, search, epidemics and cascades
- **self-similarity**, scale-free, scale-rich DW10,CM11
- dendograms, cladograms and relationship strength

Motifs

- Milo et al. first extended the concept beyond sociological networks in a 2002 article in *Science* titled: "Network Motifs: Simple building blocks of Complex Networks",
  - They defined motifs in this paper as ***patterns of interactions that occur at significantly higher rates in an actual network than in randomized networks*** and developed an algorithm for extracting them from (directed) networks



Schematic of network motif detection. Motifs are found in the real network (A) much more frequently than in a ensemble of random networks (B)

Figure by MIT OCW.

Motifs b

- Milo et al. first extended the concept in a 2002 article in *Science* titled: "Network Motifs: Simple building blocks of Complex Networks",
  - They define motifs as ***patterns of interactions that are significantly higher than in randomized networks***
  - They studied 19 networks (in six different classes)
    - For 2 gene transcription networks they found that the two different transcription systems showed the same motifs

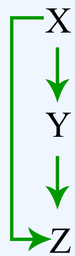
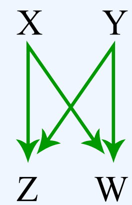
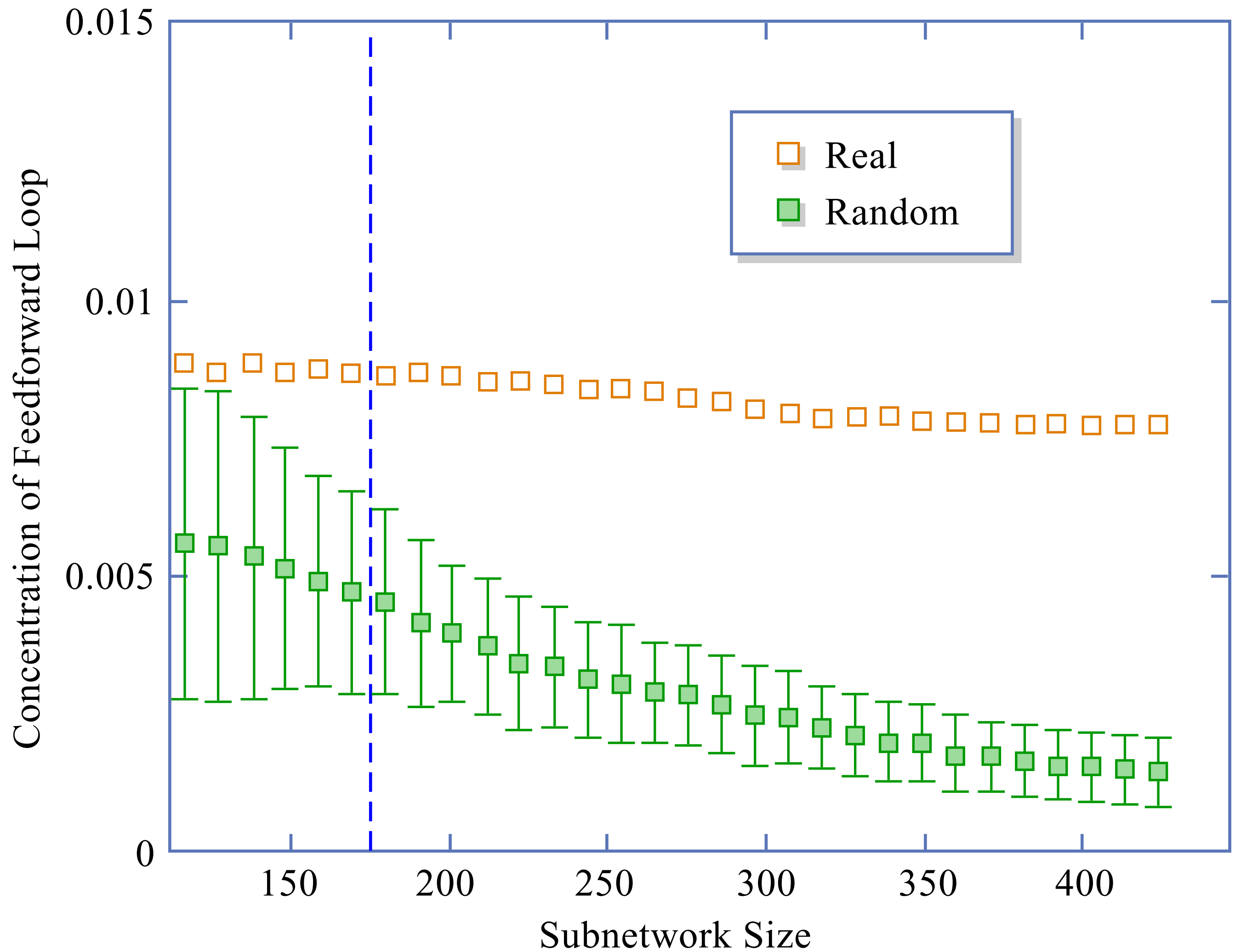
Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene Regulation (transcription)				Feedforward loop			Bi-fan				
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
S. cerevisiae*	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			

Figure by MIT OCW.

The number of times these **two motifs occur is more than 10 standard deviations greater** than their mean number of appearances in randomized networks. **None of the other 13 three node possible patterns or any other of the 199 4 node possible patterns** appear more than the **mean plus 2 standard deviations** of their appearance in randomized networks

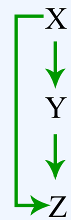
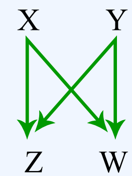
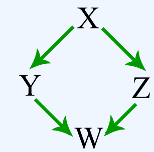
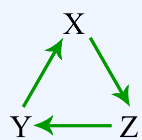
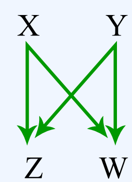
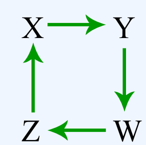


Dependence of Detectability of the 3 Node Motif on the Size of the E. coli Network Studied

Figure by MIT OCW.

Motifs c

- Milo et al. first extended the concept in a 2002 article in *Science* titled: "Network Motifs: Simple building blocks of Complex Networks",
  - They define motifs as ***patterns of interactions that are significantly higher than in randomized networks***
  - They studied 19 networks (in six different classes)
    - For 2 gene transcription networks they found that the two different transcription systems showed the same motifs
    - For 8 ***electronic circuits*** (in 2 classes), they found

Network	Nodes	Edges	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score
Electronic Circuits (forward logic chips)			 Feedforward loop			 Bi-fan			 Bi-parallel		
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic Circuits (digital fractional multipliers)			 Three-node feedback loop			 Bi-fan			 Four-node feedback loop		
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838 ‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25

The extremely high ratios for the motifs in these cases (even at small size) can probably be interpreted as evidence of design intent and for these small technological systems the importance of available modules in such systems (lecture 18) probably accounts for the reuse of the "motifs" in the variety of circuits of the same class.

Figure by MIT OCW.

The extremely high ratios for the motifs in these cases (even at small size) can probably be interpreted as evidence of design intent and for these small technological systems the importance of available modules in such systems (lecture 18) probably accounts for the reuse of the same "motifs" in the variety of circuits of the same class.

Motifs d

- Milo et al. first extended the concept in a 2002 article in *Science* titled: "Network Motifs: Simple building blocks of Complex Networks",
  - They define motifs as ***patterns of interactions that are significantly higher than in randomized networks***
  - They studied 19 networks (in six different classes)
    - For 2 gene transcription networks they found that the two different transcription systems showed the same motifs
    - For 8 electronic circuits (in 2 classes), they found reproducible motifs at high concentration for each class of circuit studied
- One interesting conclusion is that the technique can be applied to networks with variable nodes and links
- A second interesting conclusion coming from comparison of neurons, genes, food webs and electronic circuits is
- "**Information processing** seems to give rise to **significantly different structures than does energy flow**." The possible relevance to lecture 7 and past Whitney work is intriguing and addressing it would involve a research question

Motifs e

- In the software tools and example data section of the web site, you can now find “mfinder manual”
  - This entry has links to mfinder which is software (free download) for detecting motifs on networks (PC, Windows XP and Linux versions available)
  - Also comes with mDraw which allows visualization of results of mfinder.
  - Also contains network randomization methods
- Biological, electronic (and social networks) have been found to have motifs and in many cases, the motifs have been valuable in understanding such systems.
- Why might electronic and biological networks in particular show motifs? What factors or *constraints* are important in these systems?

Function on a local scale

- The motifs shown for the electronic circuits (and the biological systems) seem to show evidence of functionality imbedded within the network and pursuing a hierarchy of function within technological networks is one interesting avenue suggested by this work.
- The following slides are a brief discussion of an approach used to estimate complexity of various systems attending to hierarchy, interactions and function (Masters thesis of Pierre-Alain Martin)
- After that, we will return to looking at hierarchies of motifs on various levels which is called coarse-graining in the literature (hierarchy of function?)

Aspects of Complexity

- Number of elements (scale)
- Number of interactions
- Patterns of interactions
- Number and interaction of hierarchical levels
- Scope
  - # of functions (and their interactions)
  - # of time scales (and their interactions)
    - Feedback and diverse time delays
  - # of spatial scales (and their interactions)
- In our "network approximations", we have deliberately started with the simple end (to do otherwise risks immediate non calculability) and the question is how much complexity must be added for these to be useful in ***our systems for our purposes.***

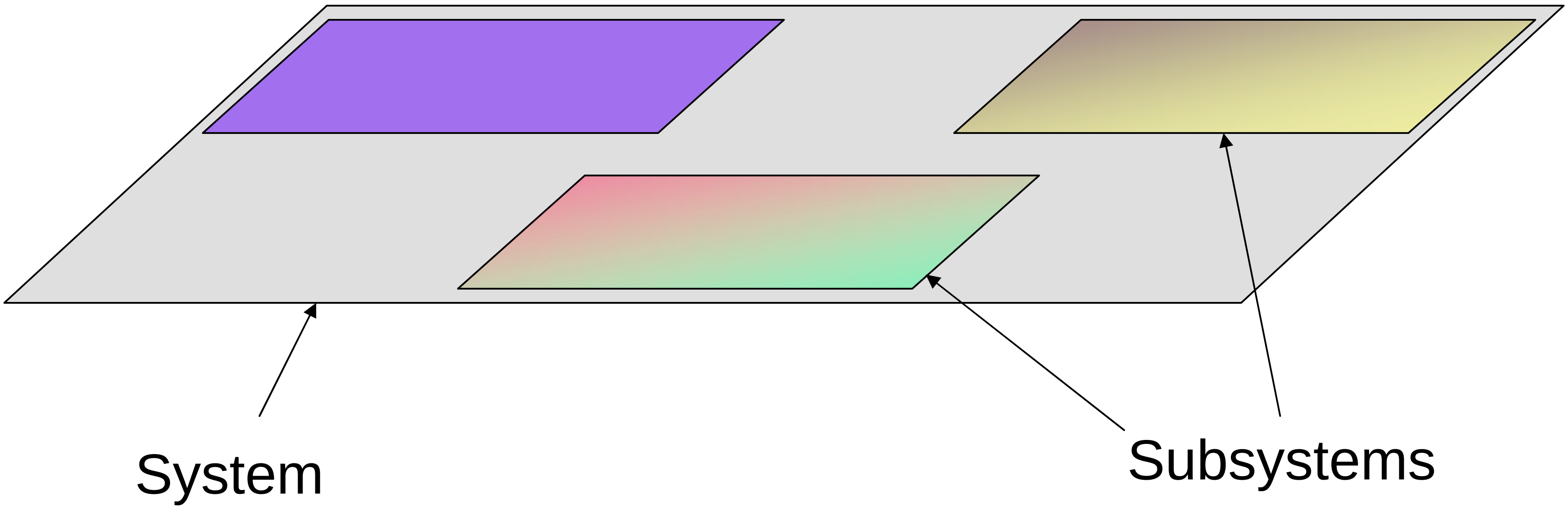
The drivers of Complexity in model and representation discussed

- Number of elements
- Number of links (JM idealization)
- Number of basic functions (new here)
- Hierarchy of these basic functions (new here)

Functional Classification Matrix

Process/Operand	Matter (M)	Energy (E)	Information (I)	Value (V)
Transform or Process (1)	**GE Polycarbonate Manufacturing Plant**	**Pilgrim Nuclear Power Plant**	**Intel Pentium V**	**N/A**
Transport or Distribute (2)	**FedEx Package Delivery**	**US Power Grid System**	**AT&T Telecommunication Network**	**Intl Banking System**
Store or House (3)	**Three Gorge Dam**	**Three Gorge Dam**	**Boston Public Library (T)**	**Banking Systems**
Exchange or Trade (4)	**eBay Trading System (T)**	**Energy Markets**	**Reuters News Agency (T)**	**NASDAQ Trading System (T)**
Control or Regulate (5)	**Health Care System of France**	**Atomic Energy Commission**	**International Standards Organization**	**US Federal Reserve (T)**

System Representation



Convention

Elements

Transform or Process

Transport or Distribute

Store or House

Exchange or Trade

Control or Regulate

1 Level 1

2 Level 2

3 Level 3

4 Level 4

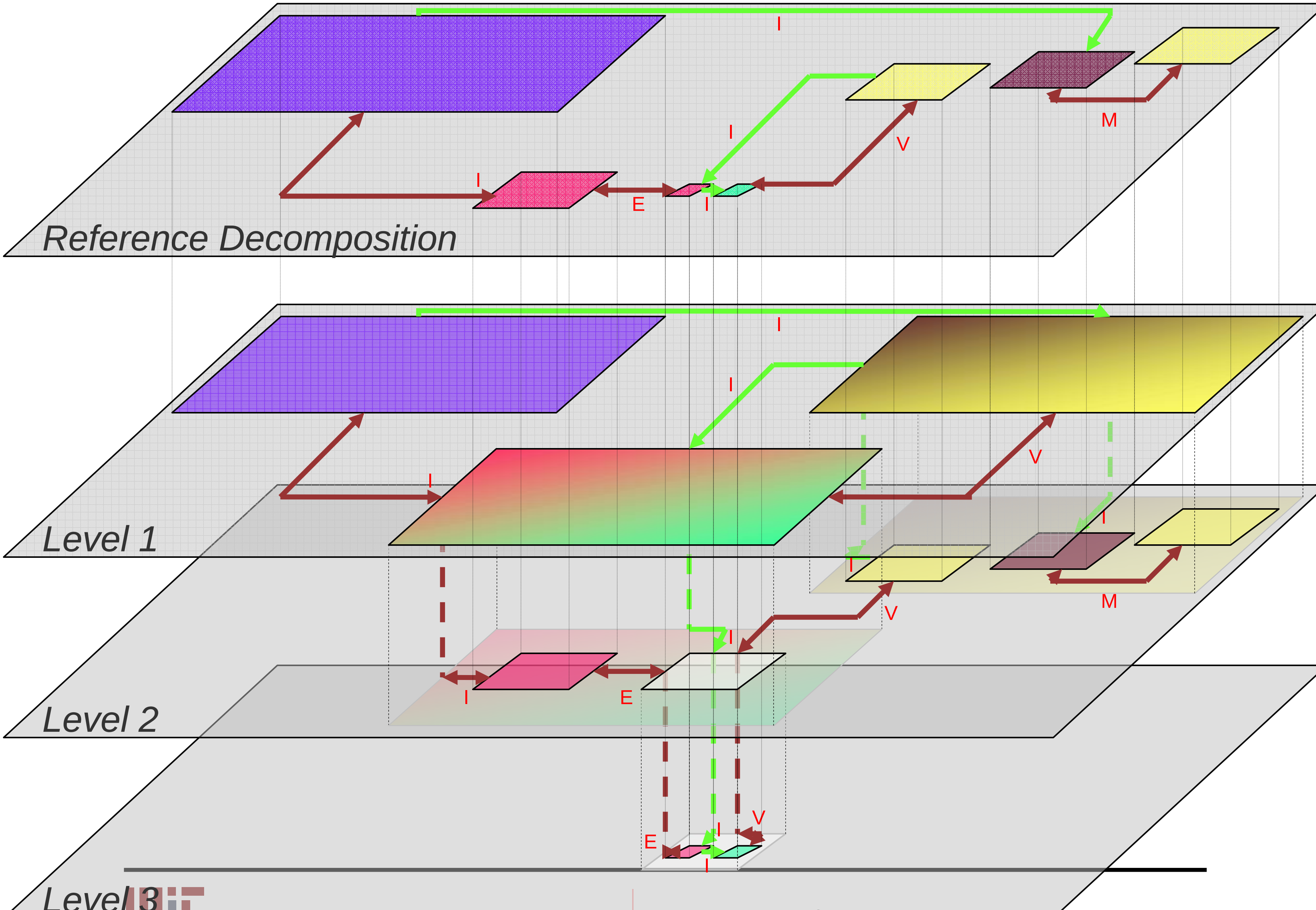
5 Level 5

Links

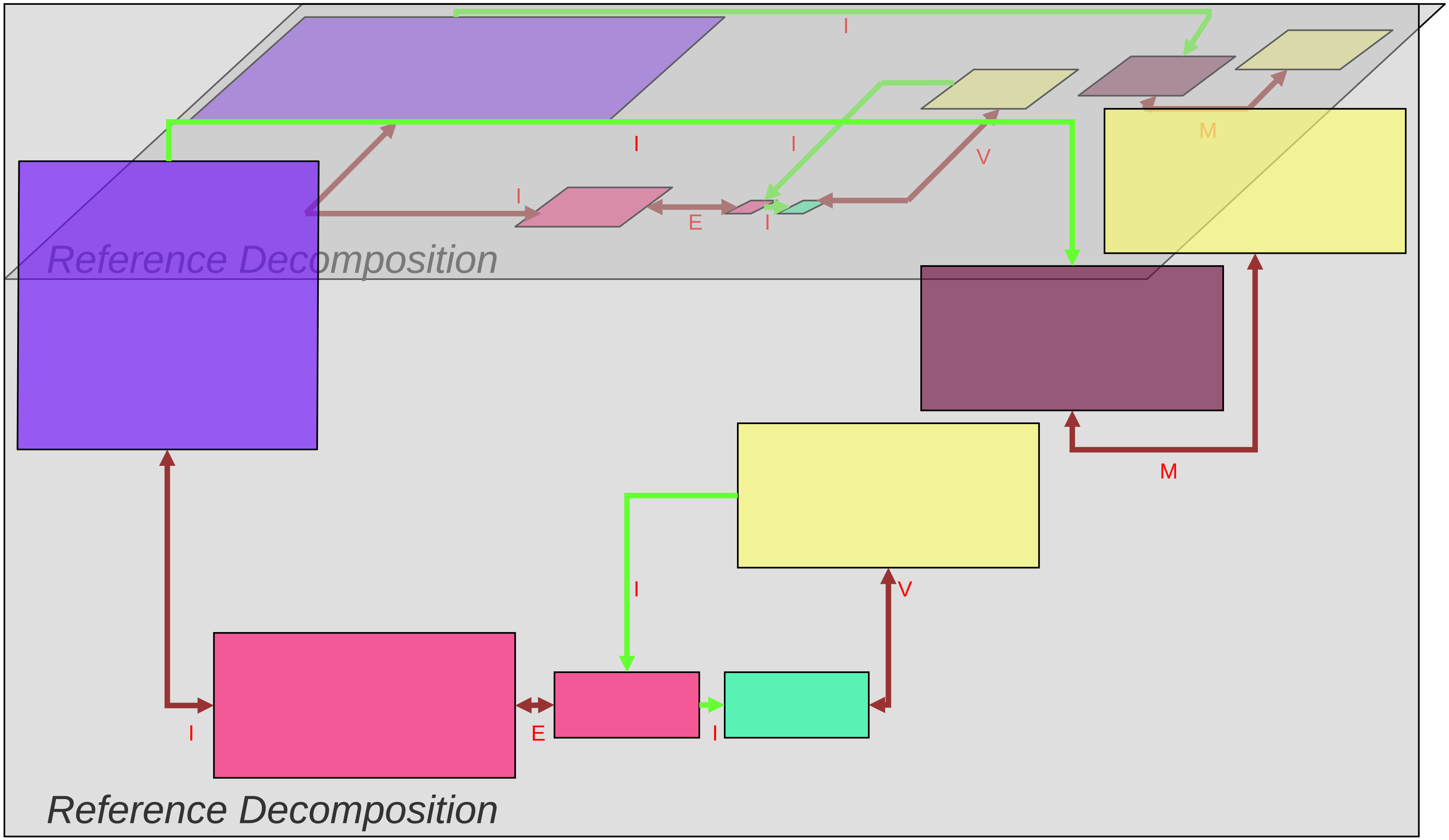
Directional

Bidirectional

I – Information
E – Energy
M – Matter
V – Value



Reference Decomposition
Level 1
Level 2
Level 3
I
E
V
M



The Reference Decomposition is a map of interconnected mono-functional elements

Recommended complexity metric

Counting only the non-identical elements

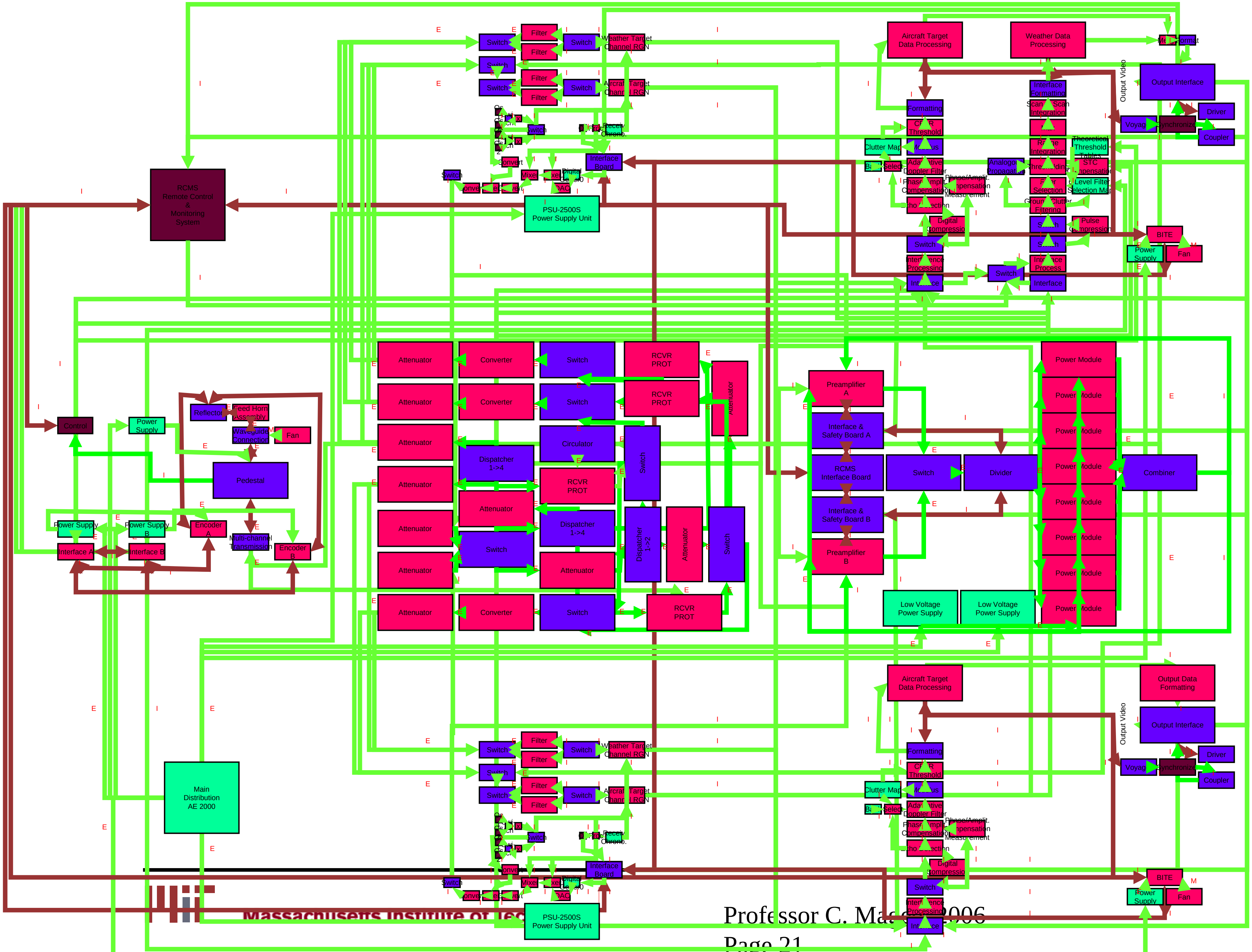
$$\mathcal{C} = \frac{L}{E} \times \sum_j E_j^{NI} \frac{\mathrm{Ln}\left((V_j + 1) \times (H_j + 1)\right)}{\mathrm{Ln}(4)}$$

Normalized Density

Product

Microscopic

V is the number of basic functions in the system
H is the number of hierarchical layers to decompose the system to monofunctional elements



RCMS
Remote Control
&
Monitoring
System
Aircraft Target
Data Processing
Weather Data
Processing
Output Interface
Output Video
Driver
Voyager
Synchronizer
Coupler
Switch
Filter
Weather Target
Channel RGN
Aircraft Target
Channel RGN
Interface
Board
PSU-2500S
Power Supply Unit
Formatting
CFAR
Threshold
Clutter Map
Adaptive
Doppler Filter
Phase/Amplit.
Compensation
Measurement
Echo Detection
Digital
Compression
Interference
Processing
Interface
Analogic
Propagation
Interface
Formatting
Scan to Scan
Integration
Range
Integration
Thresholding
Theoretical
Threshold
Tables
STC
Compensation
Level Filter
Selection Map
Filter
Selection
Ground Clutter
Filtering
Pulse
Compression
BITE
Power
Supply
Fan
Control
Reflector
Feed Horn
Assembly
Waveguide
Connection
Pedestal
Encoder
A
Encoder
B
Multi-channel
Transmission
Power Supply
Power Supply
B
Interface A
Interface B
Attenuator
Converter
RCVR
PROT
Circulator
Dispatcher
1->4
Dispatcher
1->2
Preamplifier
A
Interface &
Safety Board A
RCMS
Interface Board
Interface &
Safety Board B
Preamplifier
B
Divider
Combiner
Power Module
Low Voltage
Power Supply
Output Data
Formatting
Main
Distribution
AE 2000
Mixer
Digital
DAC
Receiver
Chrono.
Massachusetts Institute of Tec

$$\mathcal{C}\text{(STAR 2000)} = \frac{L}{E} \times \sum_j E_j^{NI} \frac{\mathrm{Ln}\,((V_j + 1) \times (H_j + 1))}{\mathrm{Ln}\,(4)} = 350.20$$

Complexity Estimation and Technological System Representation by Networks

- Martin's thesis work may be a superior way to calculate complexity and worked well for the two cases he studied.
- Much more application to other systems is needed to determine its utility.
- For today's lecture purpose, we introduce it to allow discussion of node differentiation by function and by hierarchical level. In addition, we want to note the possible utility of the representation developed in that work as a basis for developing more effective (yet tractable) network models for technological systems.

Self-similarity and self-dissimilarity

- Wolpert and Macready(2000) introduced the concept of self-dissimilarity as a complexity metric
- They defined self-dissimilarity as “the variability of interaction patterns of a system at different spatio-temporal scales”
- Wolpert and Macready invented relatively elaborate methods for statistically applying their concept and demonstrate it only through numerical simulations.
- Itzkovitz et. al (2004) have recently developed a method they call “coarse-graining” based on their prior work on motifs. This method also assesses self-dissimilarity and has been applied to biological and technological networks.

Coarse-Graining

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for "reverse-engineering" that can be applied even when the lower level functional units are unknown (biological focus).
- The coarse-grained version of a network is a new network with fewer elements. This is achieved by replacing some of the original nodes by CGU's (patterns of node interactions at the level being examined-motifs chosen somewhat differently).
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of CGU's (minimize the "vocabulary" of CGU's and the complexity of the chosen CGU's while maximizing the coverage of the original network by the coarse-grained description).

Optimal selection of CGU's

- Complexity defined (number of "ports"for a node -equivalent to JM)

$$H = I + O + 2M$$

- The number of ports in the network (system) covered by a motif group selected

$$\Delta P = P_{covered} - \sum_{i=1}^{N} n_i H_i$$

- A scoring function which can be maximized to optimize coverage and favors CGU's which have high coverage and many internal nodes (and few external mixed nodes) is

$$S = [E_{covered} + \alpha P_{covered}] - [\alpha \sum_{i=1}^{N} n_i H_i + \beta N + \gamma \sum_{i=1}^{N} T_i]($$

Coarse-Graining

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for “reverse-engineering” that can be applied even when the lower level functional units are unknown (biological focus).
- The coarse-grained version of a network is a new network with fewer elements. This is achieved by replacing some of the original nodes by CGU’s (patterns of node interactions at the level being examined-motifs chosen somewhat differently).
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of CGU’s (minimize the “vocabulary” of CGU's and the complexity of the chosen CGU’s while maximizing the coverage of the original network by the coarse-grained description).
- Applying this algorithm to an electronic circuit..

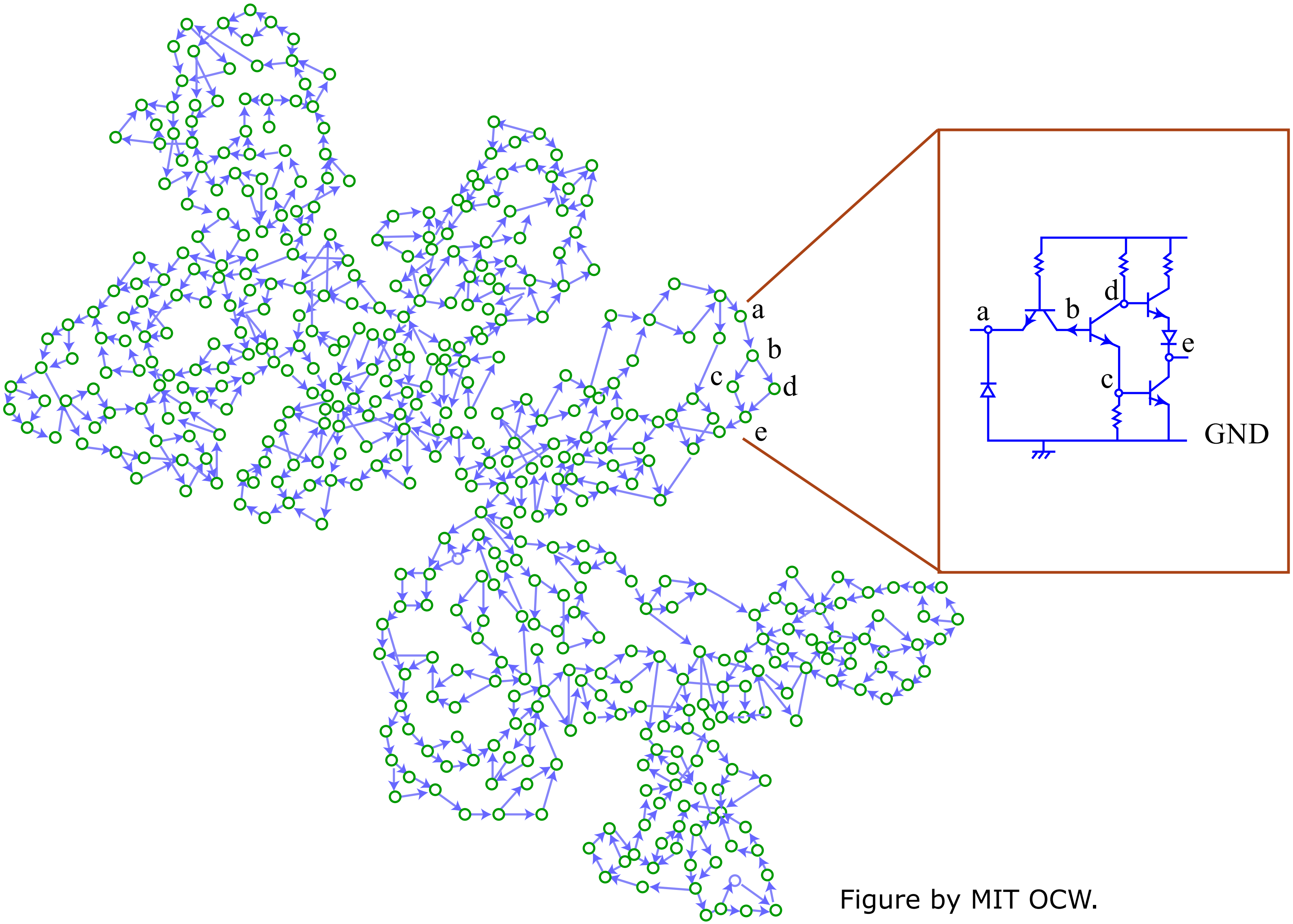


Figure by MIT OCW.

Transistor level map of an 8 bit binary counter used in a digital fractional multiplier. Highlighted is a sub-graph that represents the transistors that make up one NOT gate. Examining possible motifs **up to 6 nodes** shows…

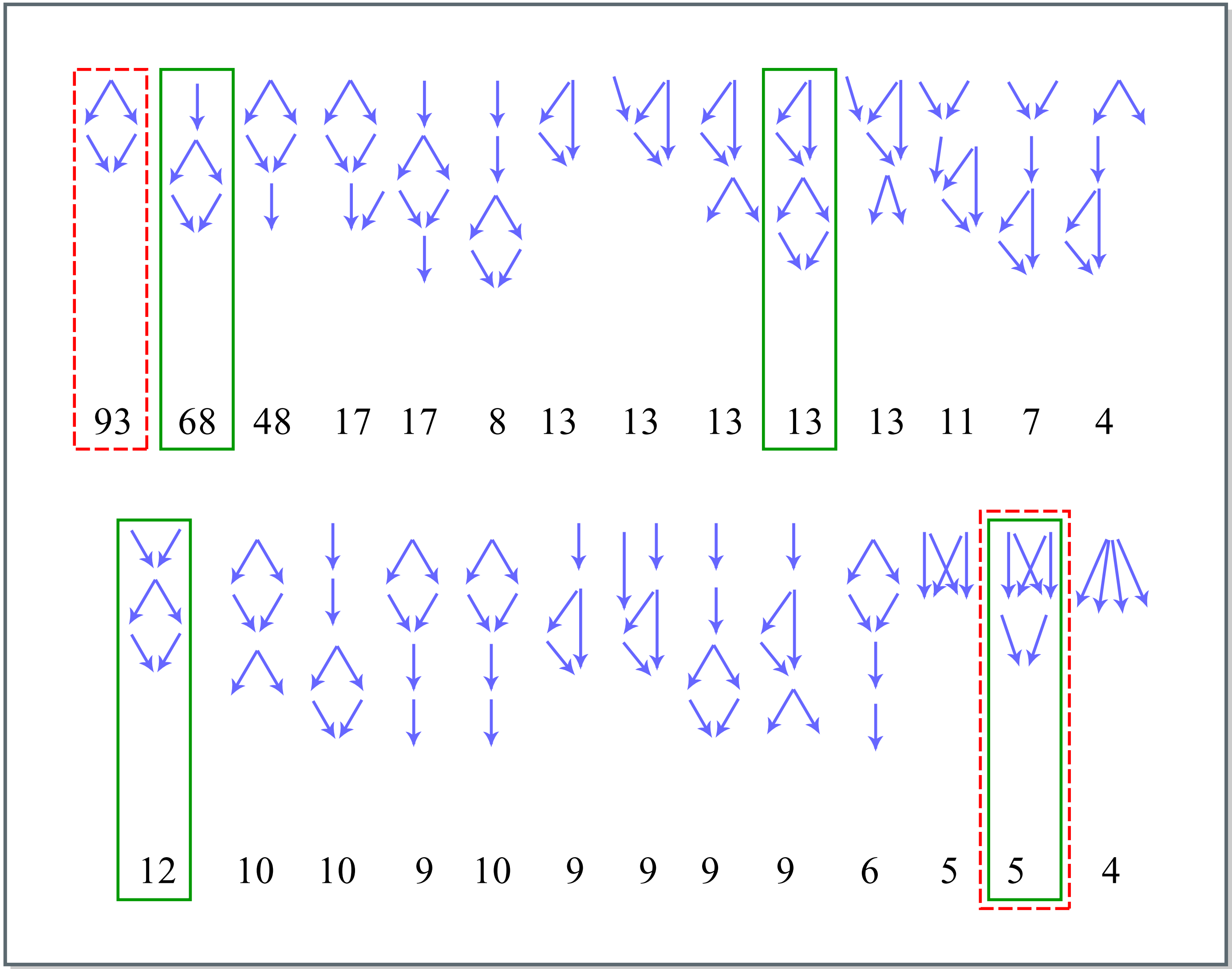


Figure by MIT OCW.

Two sets of possible optimal motif-based CGU's. The solid boxes choice can be arranged to arrive at a "gate-level coarse-graining"

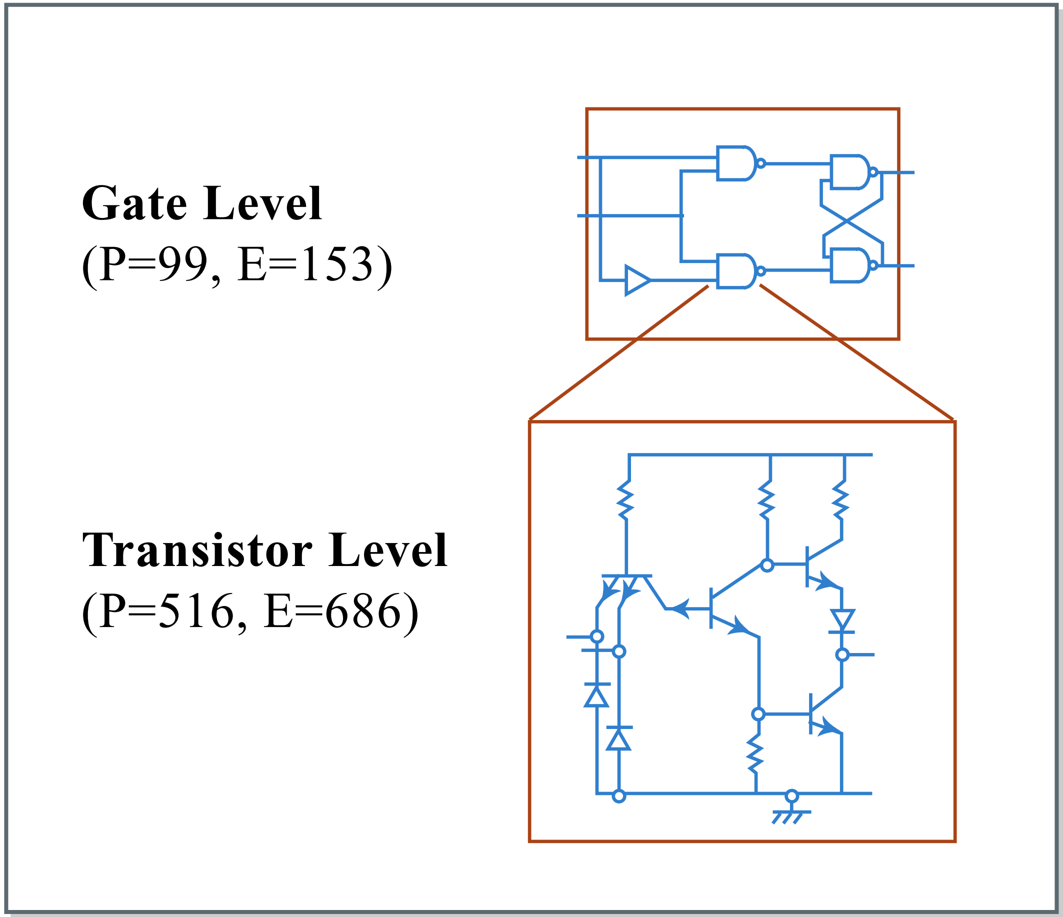


Figure by MIT OCW.

In the transistor level, nodes represent transistor junctions. In the gate level, nodes are CGU's, made of transistors, each representing a logic gate. Shown is the CGU that corresponds to a NAND gate. Re-applying the coarse-graining optimization sequentially yields 2 more levels..

Counter Level
(P=42, E=56)

Flip-Flop Level
(P=59, E=97)

Gate level
(P=99, E=153)

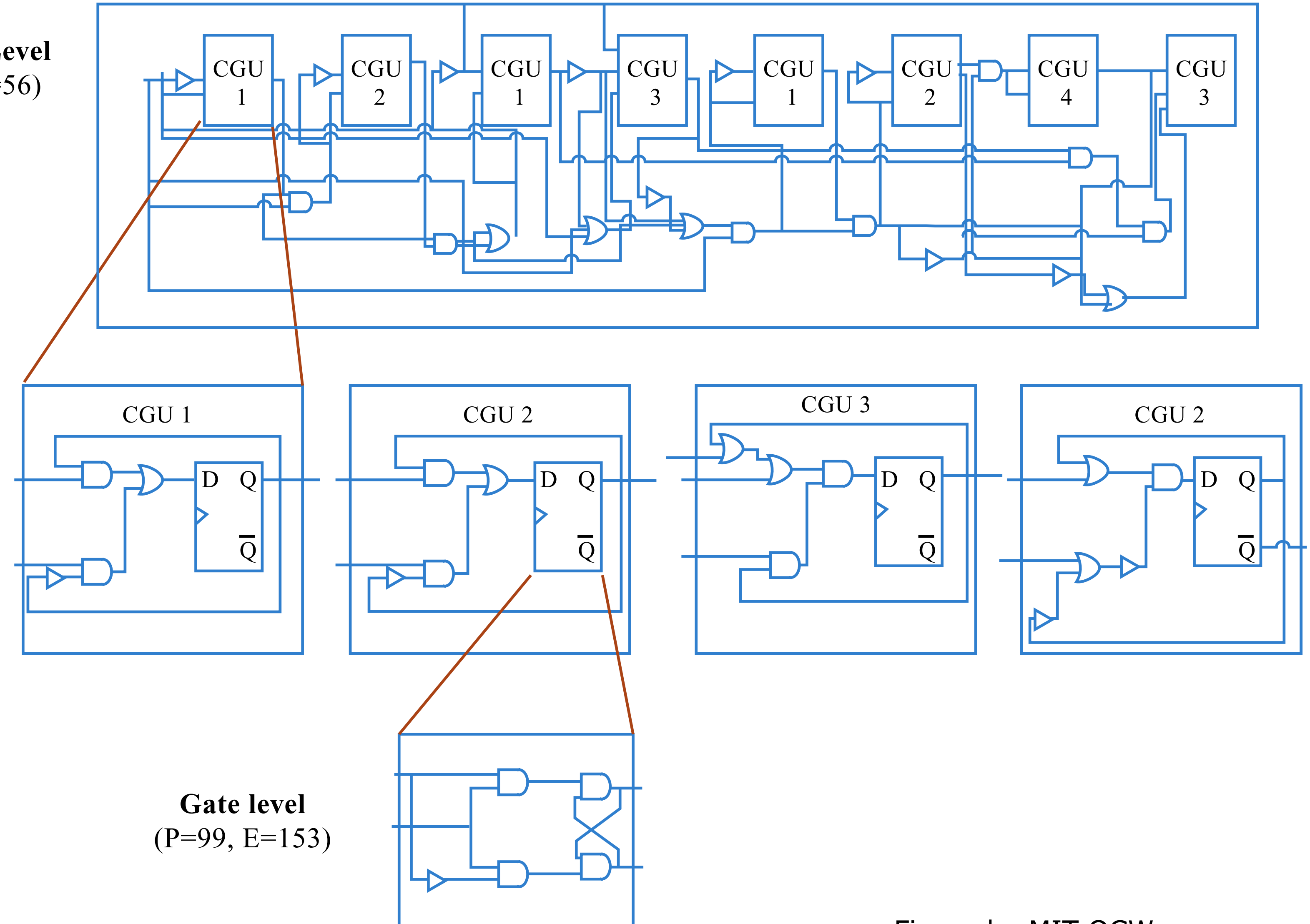
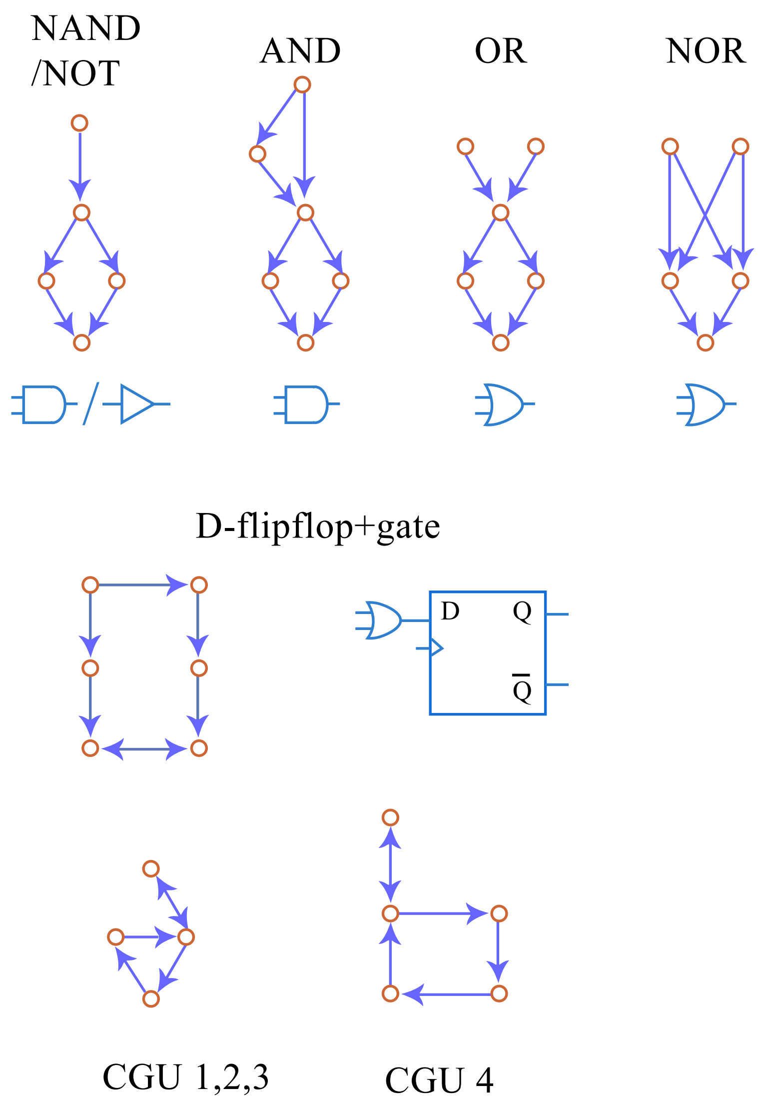


Figure by MIT OCW.

Coarse-Graining b

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for "reverse-engineering" that can be applied even when the lower level functional units are unknown (biological focus).
- The coarse-grained version of a network is a new network with fewer elements. This is achieved by replacing some of nodes by GCU's (patterns of node interactions at the level being examined.
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of GCU's (minimize the "vocabulary" of GCU's while maximizing the coverage of the original network by the coarse-grained description).
- Applying this algorithm to an electronic circuit, one finds a four level description which has variable functional significance and ***self-dissimilarity*** at each level



Gate Level

Nodes are SIUs made of transistors

Flip-Flop Level

Nodes are gates or SIUs made of gates

Counter Level

Nodes are gates or CGUs made of gates+flip-flops

Figure by MIT OCW.

Self-dissimilarity at multiple levels in the electronic circuit. This change of patterns with level apparently applies to all biological and technological networks studied thus far.

Coarse-Graining c

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for “reverse-engineering”.
- The coarse-grained version of a network is a new network with fewer elements.
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of GCU’s
- Applying this algorithm to an electronic circuit, one finds a four level description which has variable functional significance and ***self-dissimilarity*** at each level
- Note the fundamental difference between Coarse-Graining and algorithms for detection of community structure:
  - Community structure algorithms try to optimally divide networks into **sub-graphs with minimal interconnections** but these sub-graphs are distinct and complex
  - Coarse-Graining seeks a **small dictionary** of simple sub-graph types in order to elucidate the **function** of the network in terms of **recurring building blocks**

Coarse-Graining c

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for “reverse-engineering”.
- The coarse-grained version of a network is a new network with fewer elements.
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of GCU’s
- Applying this algorithm to an electronic circuit, one finds a four level description which has variable functional significance and ***self-dissimilarity*** at each level
- Note the fundamental difference between Coarse-Graining and algorithms for detection of community structure:
  - Community structure algorithms try to optimally divide networks into **sub-graphs with minimal interconnections *(modularity1***) but these sub-graphs are distinct and complex
  - Coarse-Graining seeks a **small dictionary** of simple sub-graph types in order to elucidate the **function** of the network in terms of **recurring building blocks (*modularity 2)***

Different Definitions of "Modular" or "Module" (modified from Whitney)

- You can see different elements and the places where they join (common engineering view particularly in ME but also in economics): H. Simon's near-decomposability(1962).
- Each item does a specific thing (form-function, genotype-phenotype in a one-to-one relationship) (Suh, Altenberg and also many biological papers discussing modularity)
- You need only know how to use them and don't need to know what's inside (common engineering view particularly in EE and software)
- Interconnectedness is concentrated inside them (Alexander, software design)
- Their links to the outside are standardized , or simple and few (Alexander)
- Modules can be replaced in a system arbitrarily preserving (but possibly modifying) function: (Plug and Play intuition)

Different Definitions of “Modular” or “Module”

- You can see different elements and the places where they join **(modularity 1)**
- Each item does a specific thing (form-function, genotype-phenotype in a one-to-one relationship) (Suh, Altenberg) **(modularity 2)**
- You need only know how to use them and don’t need to know what’s inside **(modularity 2)**
- Interconnectedness is concentrated inside them (Alexander)(software design) **(modularity 1)**
- Their links to the outside are standardized **(modularity 2**), or simple and few (Alexander) **(modularity 1)**

"Better" Definition(s) of Modularity

- Modularity 1:
  - The system can be decomposed into subunits to arbitrary depth
  - These subunits can be dealt with separately (to some degree)
    - In different domains, such as design, manufacturing, use, error-correction, recycling
- Modularity 2:
  - The functions of the system can be associated with clusters of physical elements
    - in the limit one function:one module
  - These elements operate (somewhat) independently
  - (They do not have to be physically contiguous)
- Merged definition
  - The intuitive "Plug and Play" requires both definitions to be operable (without qualifications in brackets)
- There are physical constraints in moderately higher power systems that prevent modularity without significant qualification (Whitney)

Coarse-Graining d

- Itzkovitz et. al. investigate Coarse-Graining as an objective means for "reverse-engineering".
- The coarse-grained version of a network is a new network with fewer elements.
- Itzkovitz et. al. apply simulated annealing to arrive at an optimum set of GCU's
- Applying this algorithm to an electronic circuit, one finds a four level description which has variable functional significance and ***self-dissimilarity*** at each level
- Note the fundamental difference between Coarse-Graining and algorithms for detection of community structure
- Note that motifs and coarse-graining have thus far only been applied to fairly simple technological systems
    - Monofunctional from the Martin-Magee perspective and easily functionally modularized

Research Questions

- Should we apply community structure separation and coarse-graining at different levels for improved understanding (and design?) of complex technological (and biological) systems?
- Does the form-function relationship only work for relatively simple systems?
- To what degree, do the two kinds of modularity apply to different levels of abstraction and/or power/information differences?
- Hypotheses:
  - For a truly complex system, one has multiple functions that cannot be separately decomposed.
  - For such a system, one might decompose (sequentially) by community structure (interaction density) to the level of mono-functionality (Martin-Magee representation arrived at objectively)
  - One then could examine the resulting mono-functional systems by coarse-graining looking for more basic form-function relationships

Self-similarity and self-dissimilarity b

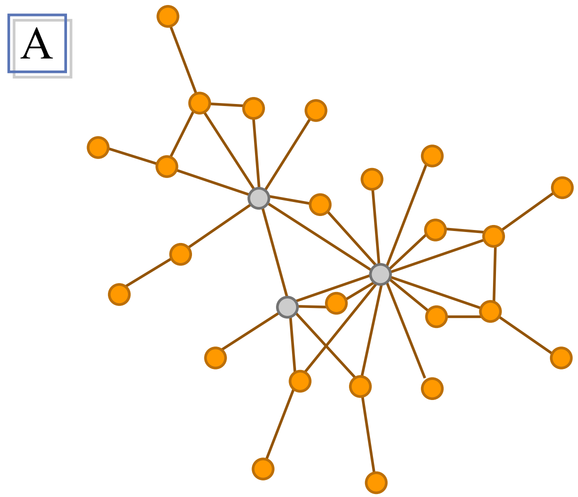
- Wolpert and Macready(2000) introduced the concept of self-dissimilarity as a complexity metric
- Self-dissimilarity is defined as "the variability of interaction patterns of a system at different spatio-temporal scales"
- *Note that as defined this definition is in a sense counter to the notion (often loosely defined) of "scale free" which implies (at least seems to) the notion that structure is repetitive at various scales*
- Wolpert and Macready invented relatively elaborate methods for statistically applying their concept and demonstrate it only through numerical simulations
- Itzkovitz et. al (2004) have recently developed a method they call "coarse-graining" based on their prior work on motifs. This method also assesses self-dissimilarity and has been applied to biological and technological networks.

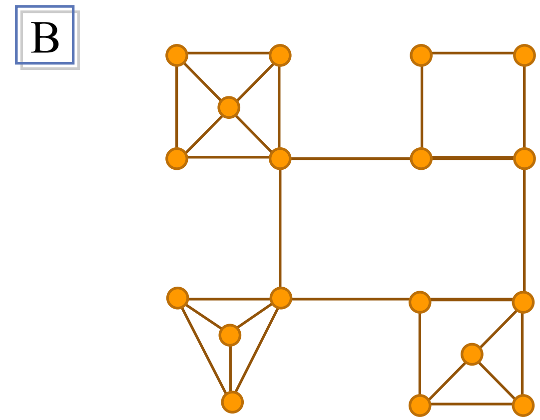
Hierarchical Organization of Modularity in Metabolic Networks

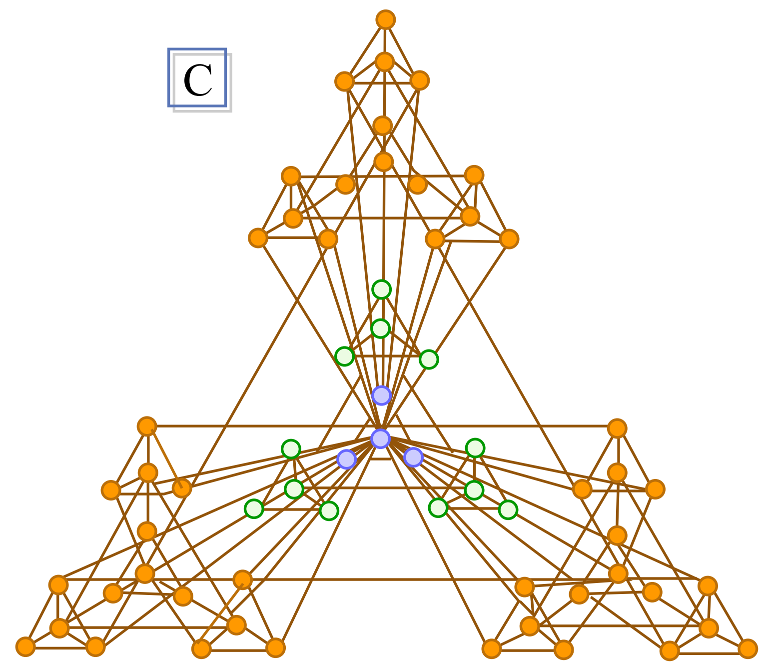
E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Baraba´si

SCIENCE VOL 297

30 AUGUST 2002 p 1551







A. Scale free
B. Modular, not scale free
C. Nested modular, scale free

Probabilistic Generation of Network Motifs (PGNMs)

- Based on idealized block models arrived at by assigning roles to nodes (limited number of roles) and defined relationships between nodes of differing roles (types). A modified scoring function is used

$$S = E_{covered} + \alpha \Delta P - \beta N - \gamma \sum_{i=1}^{N} T_i - \delta \sum_{i \in \{CGU_g\}}^{N} d_i$$

- Where g includes generalizations of the CGU's according to the block model idealization (an example of generalizations relative to the BiFan motif is shown in the next slide)

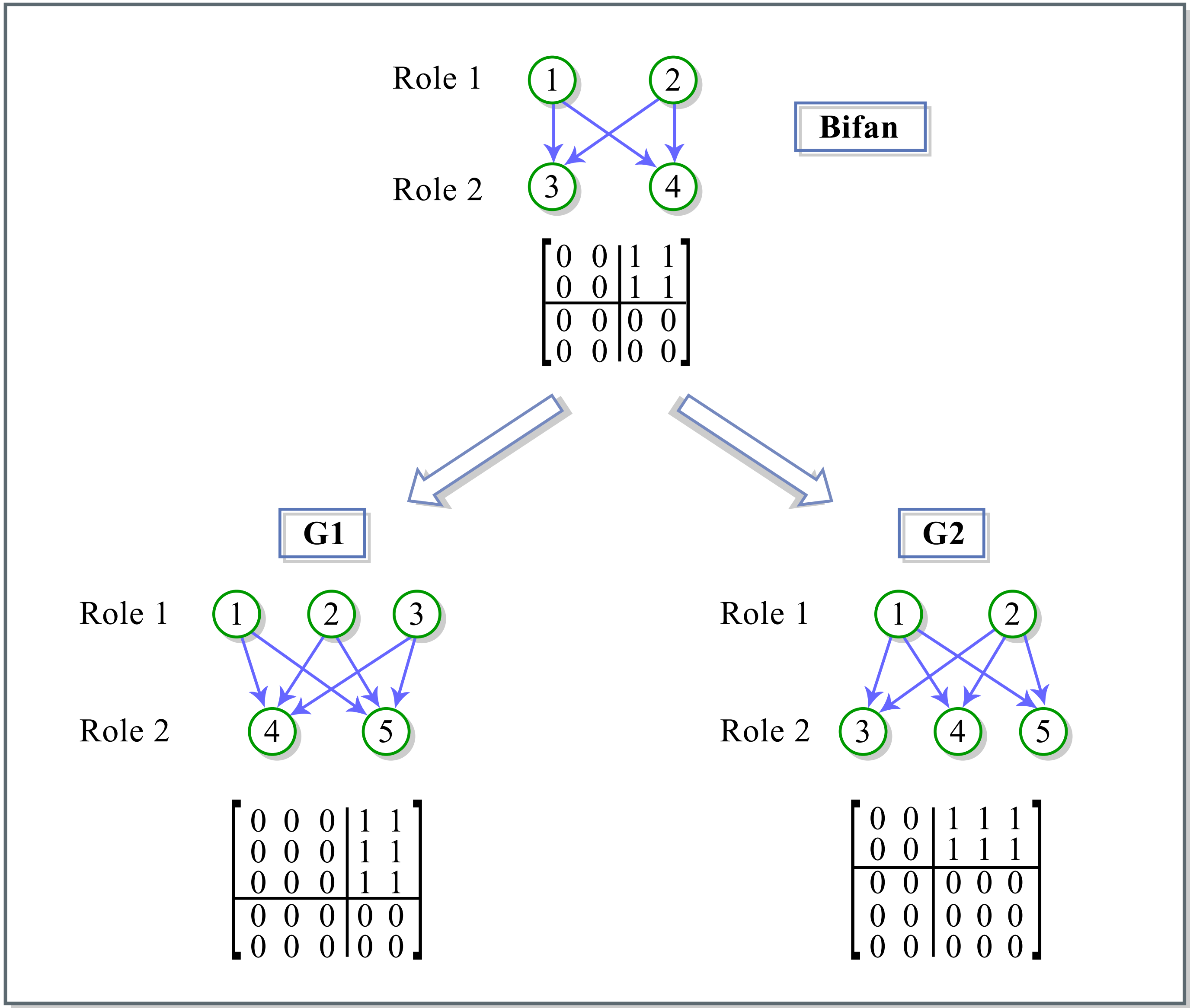


Figure by MIT OCW.

Comparison of this idealization to two cases is shown in the following slide

Probabilistic Generation of Network Motifs (PGNMs) II

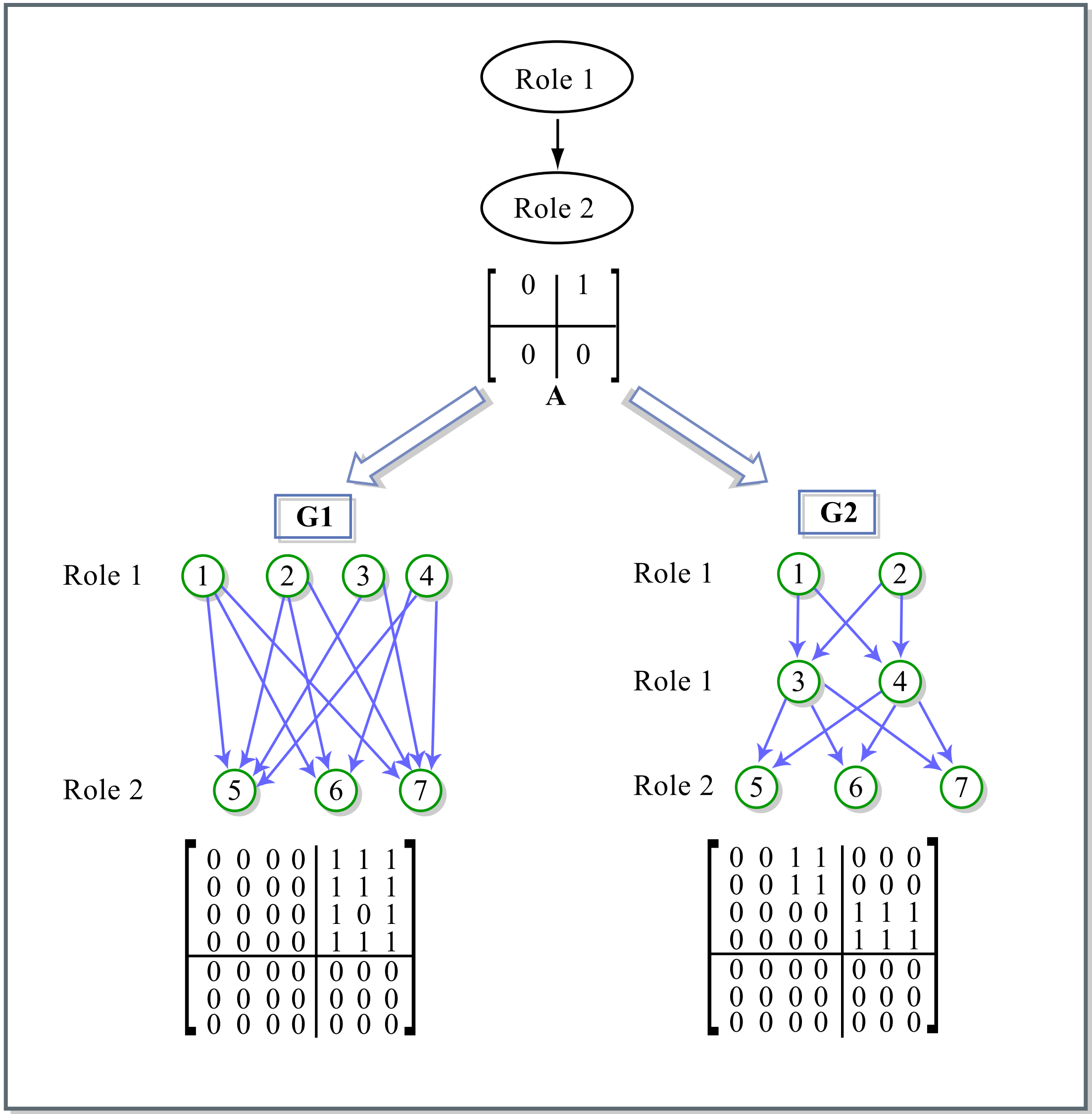


Figure by MIT OCW.

Self-similarity and self-dissimilarity c

- The scale-free modular example is self-similar.
- However, the electronic circuits and biological systems studied by Itzkovitz et. al are not scale free in that even though the modules are consistent with one another at a given scale, the patterns are dissimilar (in the Wolpert/Macready sense) at different scales (or levels of agglomeration)
- Note that the electronic circuit systems shown to be “scale-rich” by Itzkovitz et. al. show power law degree relationships so the use of the term “scale-free” when power laws is observed is nonsensical. The lack of correlation between structure and power laws was mentioned in lecture 6.
- Li et. al and Doyle et al have introduced the phrase “scale-rich” partly in response to the work by Itzkovitz and have developed some other metrics (related to degree correlation, r) and we will return to this theme in a later discussion of models of the Internet

References

- 1. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, “Network Motifs: Simple building blocks of Complex Networks”, ***Science***, Vol. 298, pp 824-827, (2002)
- 2. Pierre-Alain Martin, “A Framework for Quantifying Complexity and Understanding its Sources: Application to two Large-Scale Systems” SM thesis, MIT, 2004.
- 3. D. H. Wolpert and W. Macready, “Self-dissimilarity: An empirically observable complexity Metric”, Unifying themes in complex systems, New England Complex Systems Institute(2000). A second paper found on the NASA Moffet web site has similar information and is titled “Self-dissimilarity as a high dimension complexity measure” (2004?)
- 4. S. Itzkovitz, R. Levitt, N. Kashtan, R. Milo, M. Itzkovitz and U. Alon, “Coarse-Graining and Self-Dissimilarity of Complex Networks”, (Oct. 2004)