Motif of Sequence

Motif in Sequence

Shin-Kap Han

Seoul National University

“Sequence”

- Array of elements
  - Regularities in connections
    - Narrative order
    - Sequential (temporal) dependency
    - Interlocked contingencies
    - Processes of unfolding
    - Mechanisms of entailment
    - Structures of temporal space

- Looking for common sequence patterns:
  - Global multiple alignment
  - Local multiple alignment
    - Aims to locate relatively short patterns shared by otherwise dissimilar sequences

“Motif”

mo·tif *noun* \mō-ˈtēf\

Definition of MOTIF

1 : a usually recurring salient thematic element (as in the arts); *especially* : a dominant idea or central theme

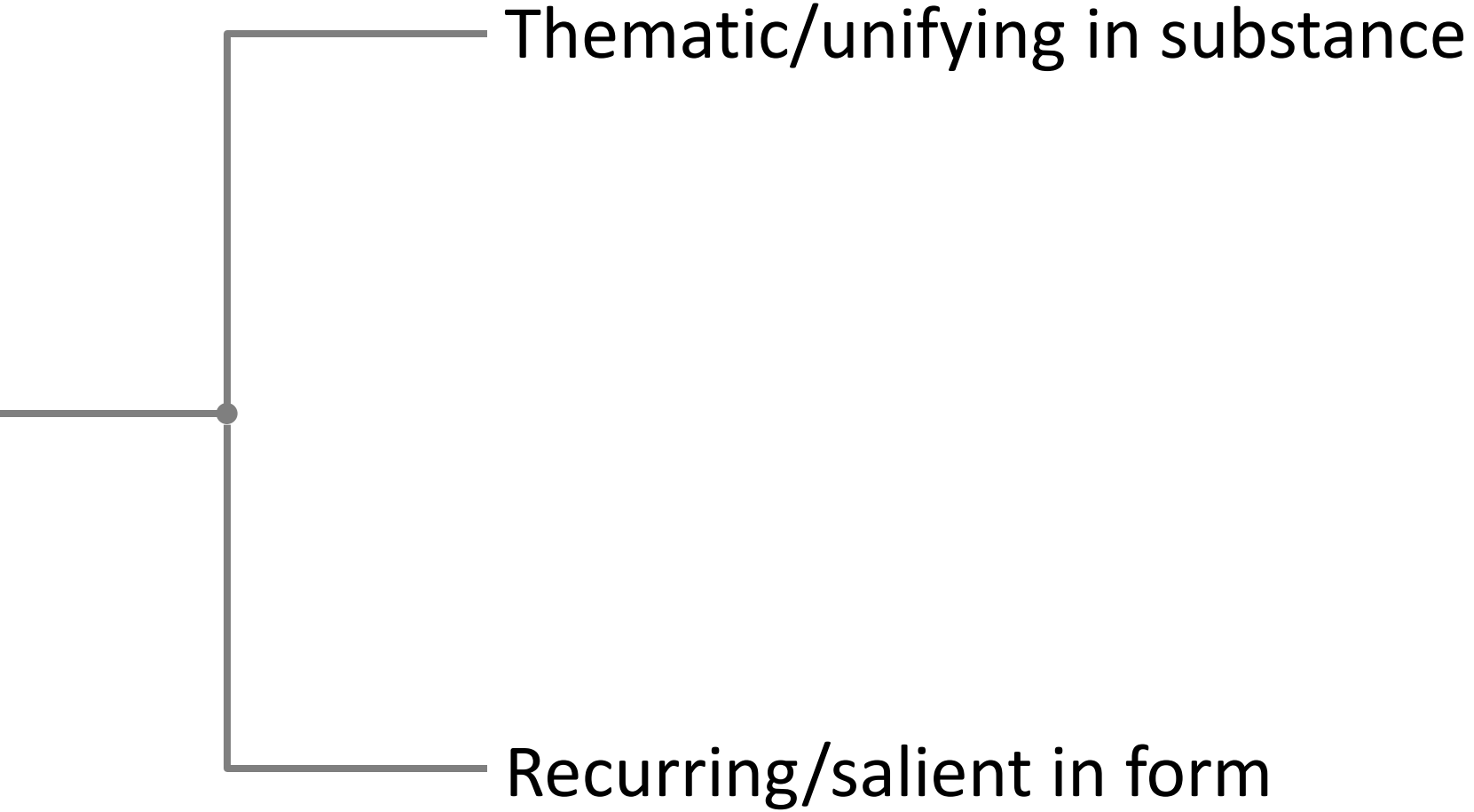
2 : a single or repeated design or color

— **mo·tif·ic** *adjective*

See motif defined for English-language learners »
See motif defined for kids »

Examples of MOTIF

- The wallpaper has a flower *motif*.
- <the *motif* of mute figures standing in lonely isolation is a recurrent one in the artist's works>
- … a hip awareness of its own cheesy implausibility, right down to the music: The thunderously orchestrated score uses "Itsy Bitsy Spider" as a *motif*. —*People*, 29 July 2002
- In retrospect, it is now clear that the alien invasion *motif* in 1950s science fiction movies reflected the Cold War atmosphere of the period. —Paul A. Cantor, *Gilligan Unbound*, 2001
- The first-class scowl, shaved head and scars on his right shoulder and biceps fit the tough-guy *motif*, but it's a facade. —Ric Bucher, *ESPN*, 28 May 2001
- The branding is done by combining a commercial trademark with one or another subcultural *motif*, a subculture the buyer belongs to or wants to join: surfing, skateboarding, … —John Seabrook, *New Yorker*, 20 Sept. 1999

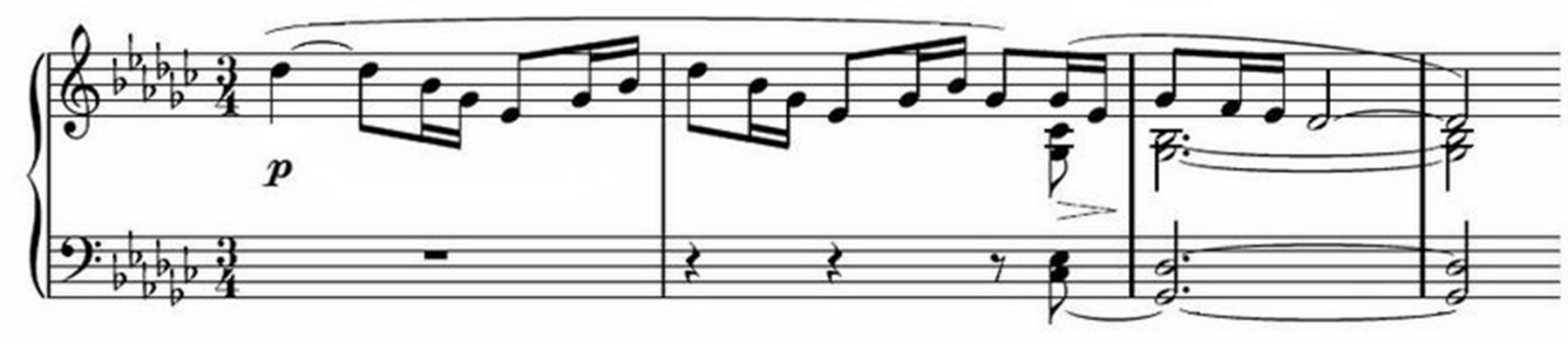

Thematic/unifying in substance
Recurring/salient in form

Face the Music

La Fille aux Cheveux de Lin

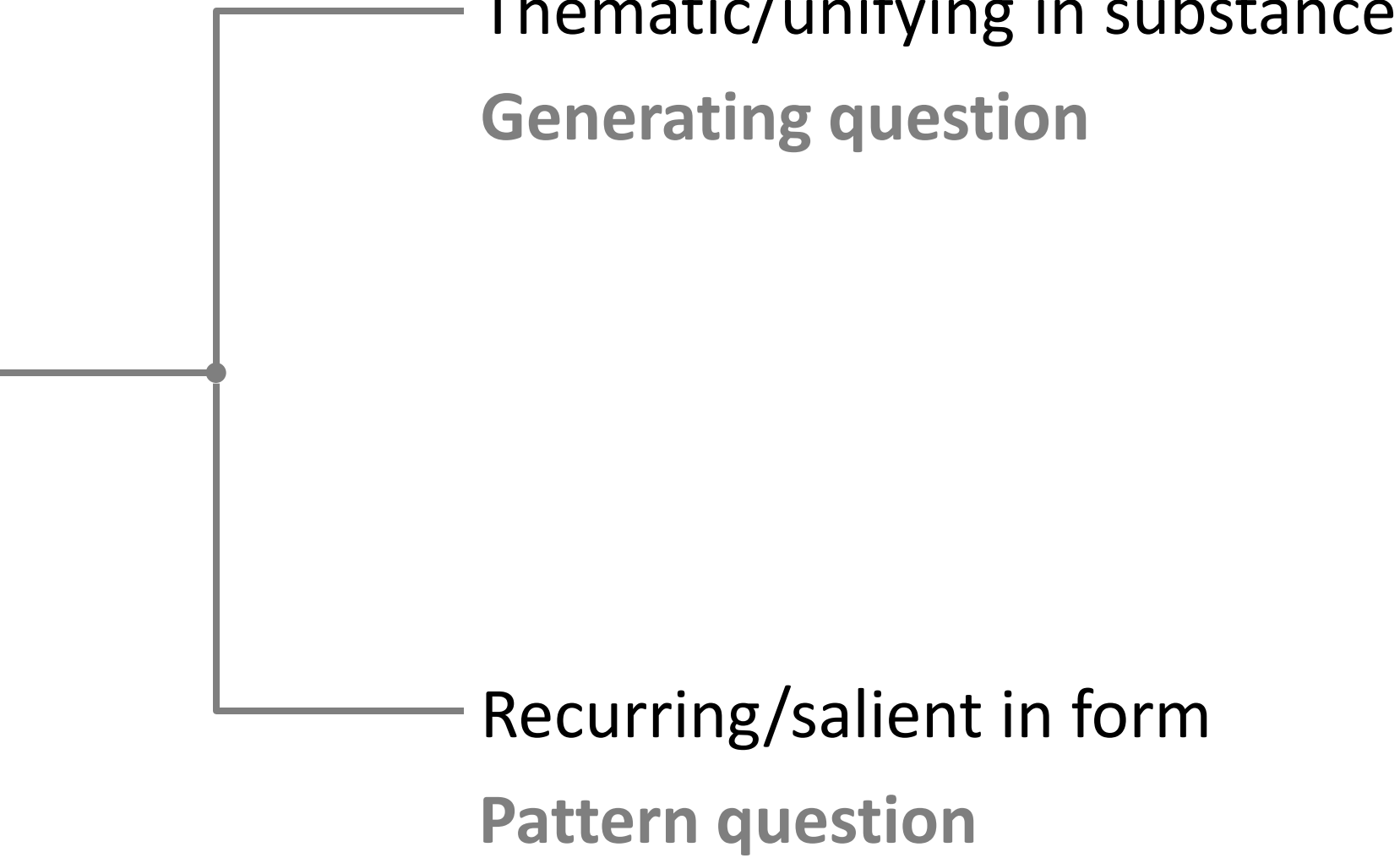
from Préludes, Book 1, No. 8

Claude Debussy



the wheat

the chaff


Thematic/unifying in substance
Generating question
Recurring/salient in form
Pattern question

Clocking Out: Temporal Patterning of Retirement

Shin-Kap Han and Phyllis Moen
Cornell University

0002-9602/2000/10501-0006$02.50

Clocking Out: Temporal Patterning of Retirement

Shin-Kap Han and Phyllis Moen
Cornell University

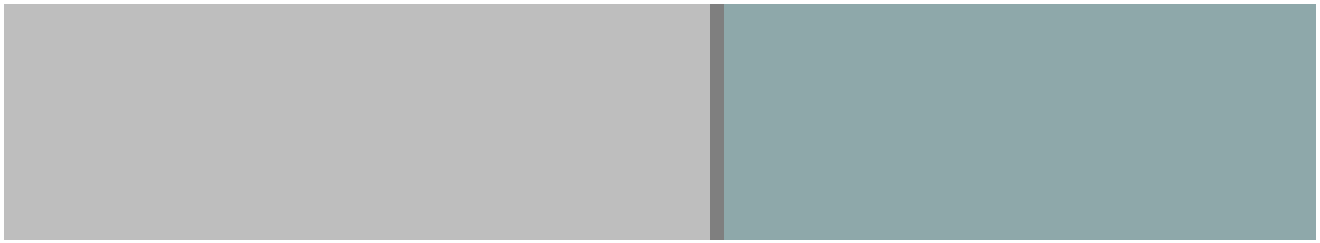
Imprinting

0002-9602/2000/10501-0006$02.50

AJS Volume 105 Number 1 (July 1999): 191–236

Clocking Out: Temporal Patterning of Retirement

Shin-Kap Han and Phyllis Moen
Cornell University

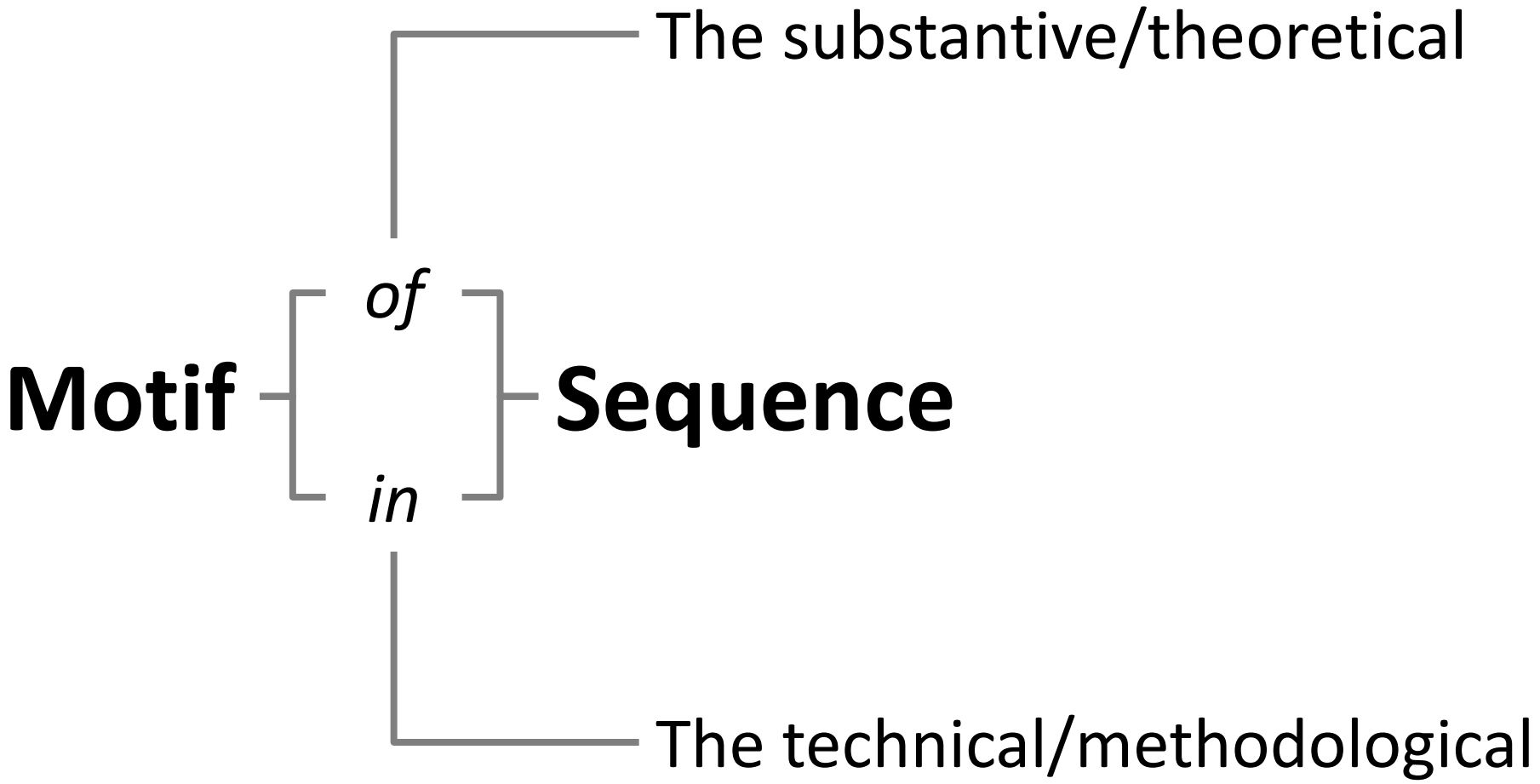


Turning point

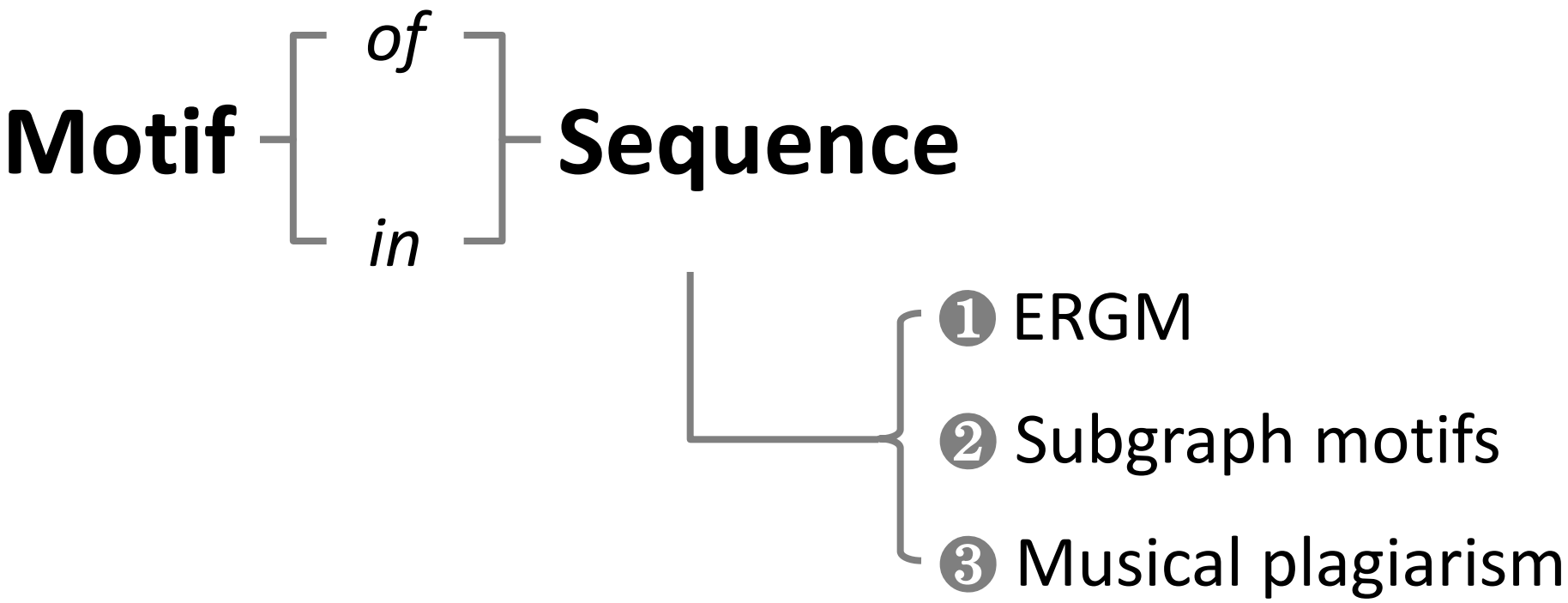
0002-9602/2000/10501-0006$02.50

AJS Volume 105 Number 1 (July 1999): 191–236

Motif as a subsequence


The substantive/theoretical
of
Motif
Sequence
in
The technical/methodological

Resources to Mobilize

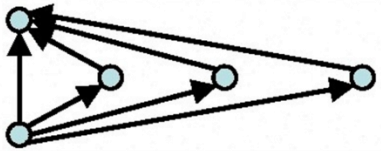
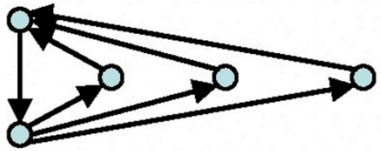
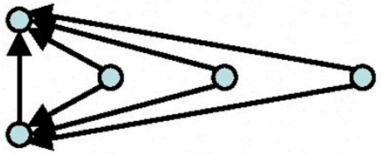
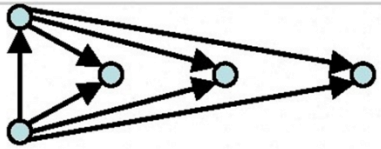
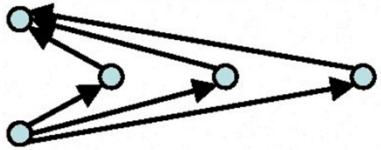
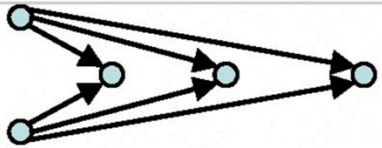
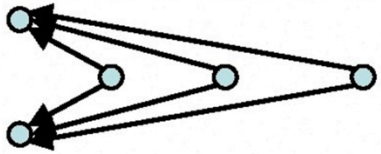


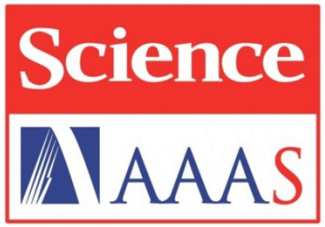
Exponential Random Graph Model

- Basic model

$$P(Y = y) = exp\left\{\sum_{k=1}^{K} \theta_k g_k(y)\right\} / \kappa(\theta)$$

$$P(Y) \propto \theta_1 g_1(y) + \theta_2 g_2(y) + \cdots + \theta_k g_k(y)$$

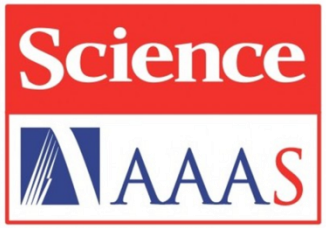
AT-T		AT-C	
AT-D		AT-U	
A2P-T		A2P-U	
A2P-D			



Detecting Subtle Sequence Signals: A Gibbs Sampling Strategy for Multiple Alignment

CE Lawrence, SF Altschul, MS Boguski, JS Liu, AF Neuwald, JC Wootton
Vol. 262 no. 5131 (8 October 1993) pp. 208-214

- Patterns shared by multiple protein or nucleic acid sequences shed light on molecular structure, function, and evolution.



Network Motifs: Simple Building Blocks of Complex Networks
R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, U Alon
Vol. 298 no. 5594 (25 October 2002) pp. 824-827

- To uncover their structural design principles, we defined “network motifs,” patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks.
- This approach may uncover the basic building blocks of most networks.

Network Motifs

- A small set of characteristic patterns (subgraphs) that occur much more frequently than in randomized networks with the same degree sequence.
- Network motifs (e.g., FFL) were demonstrated to play key information processing roles in biological regulation networks.

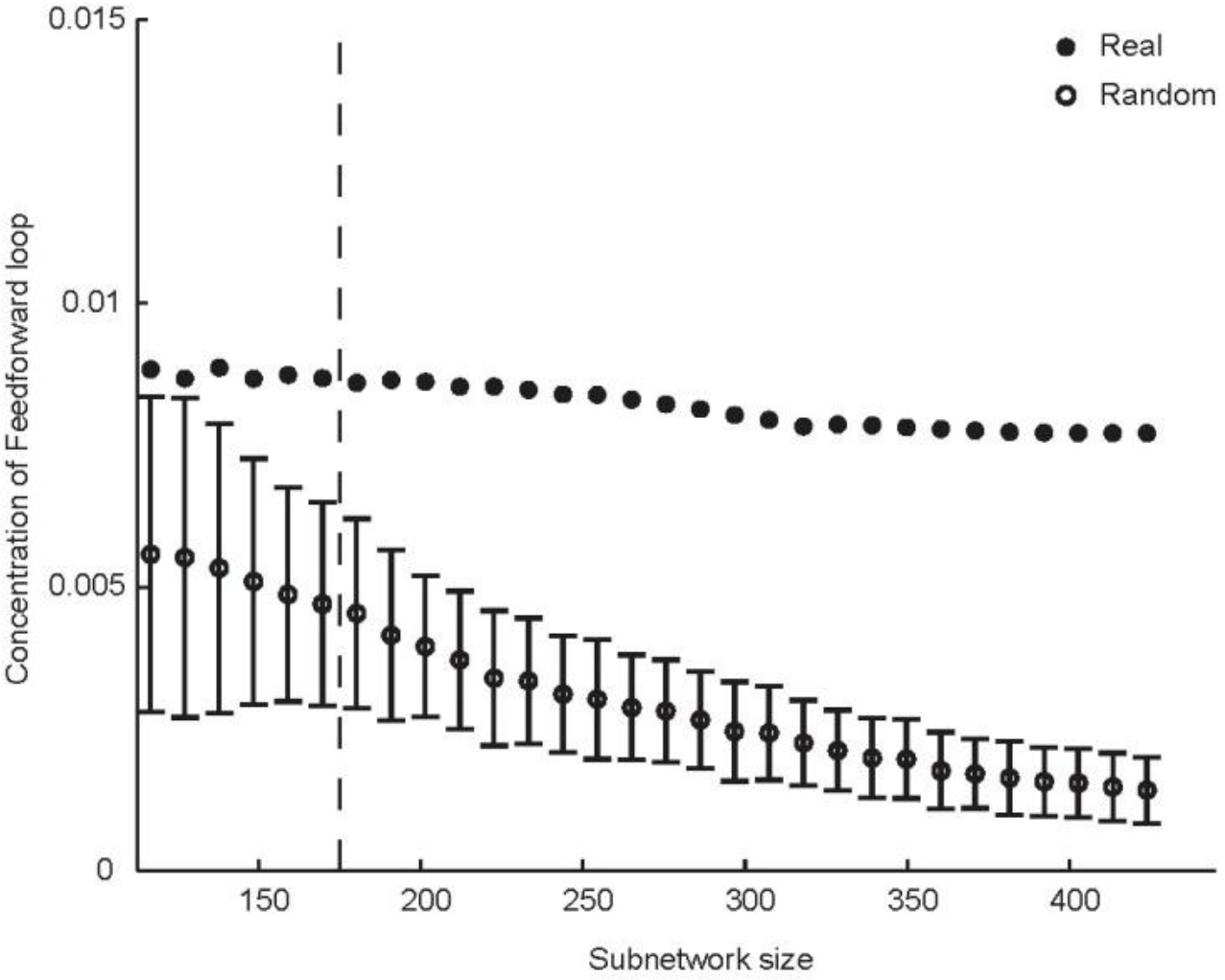
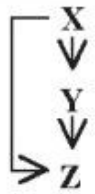
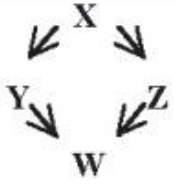
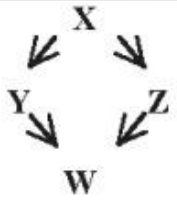
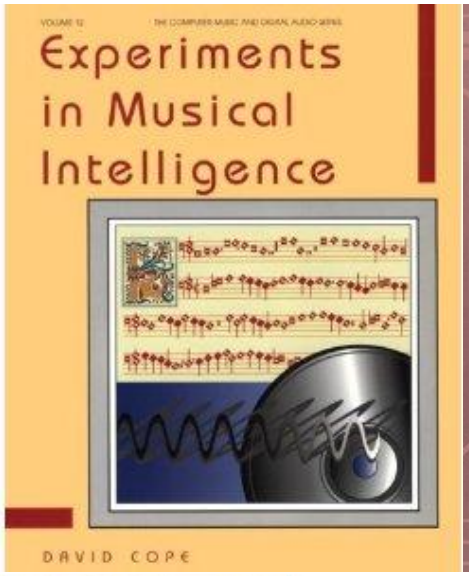


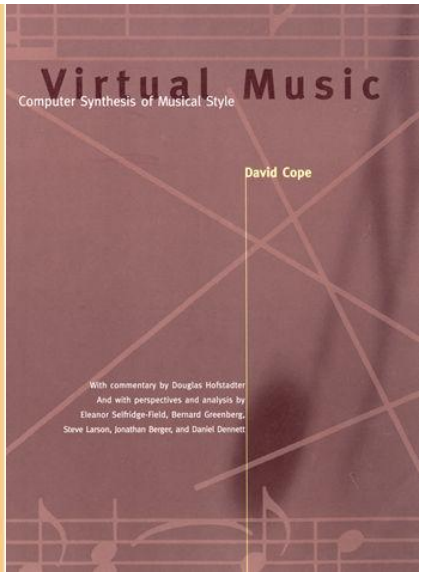
Fig. 3. Concentration C of the feedforward loop motif in real and randomized subnetworks of the *E. coli* transcription network (*11*). C is the number of appearances of the motif divided by the total number of appearances of all connected three-node subgraphs (Fig. 1B). Subnetworks of size S were generated by choosing a node at random and adding to it nodes connected by an incoming or outgoing edge, until S nodes were obtained, and then including all of the edges between these S nodes present in the full network. Each of the subnetworks was randomized (*17*, *18*) (shown are mean and SD of 400 subnetworks of each size).

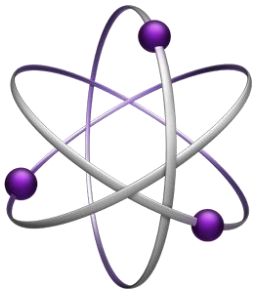
Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			

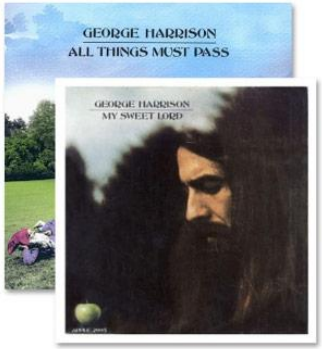
Experiments in musical intelligence

Idiosyncratic court findings

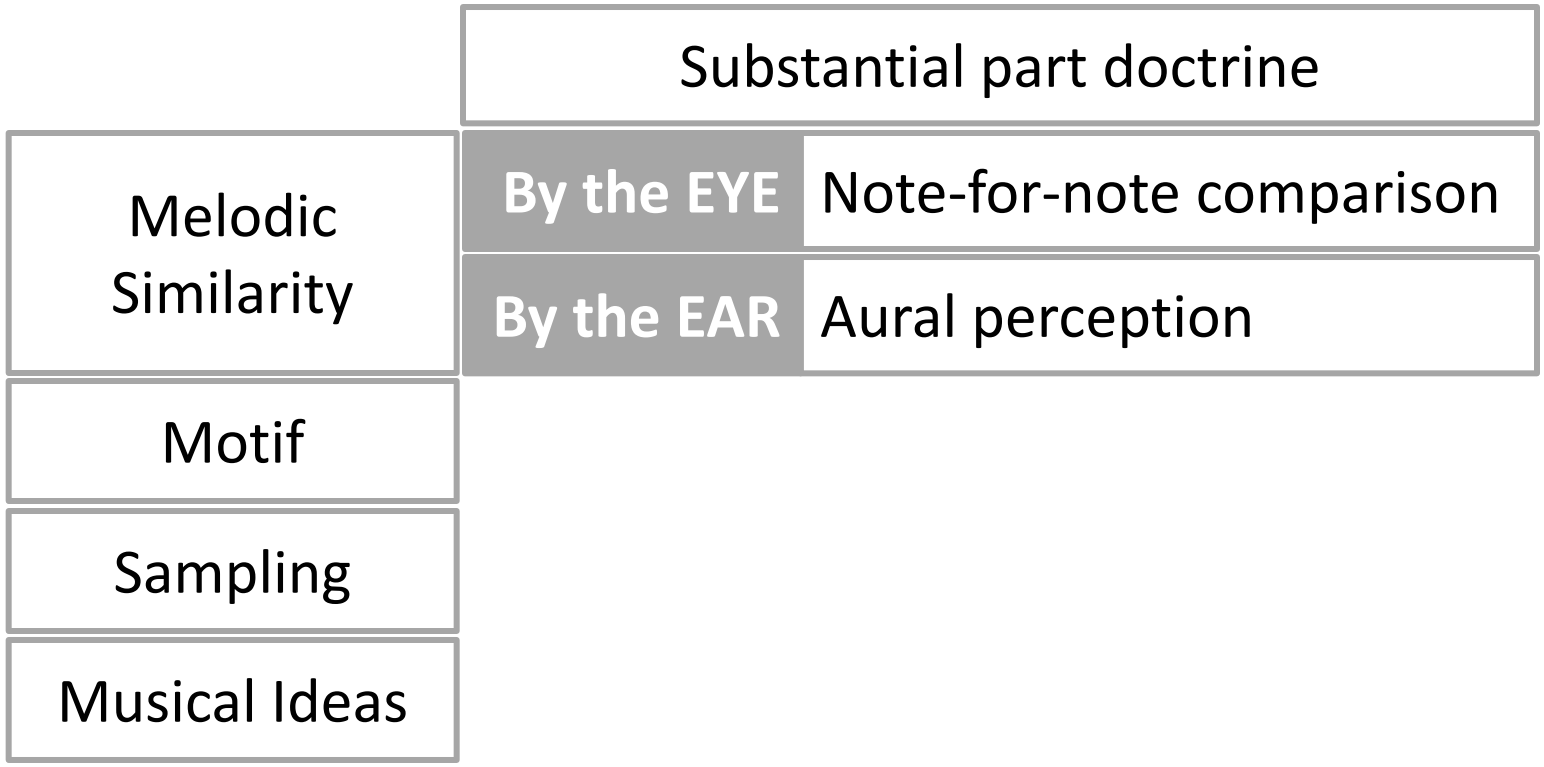








Musical Plagiarism



Musical Plagiarism

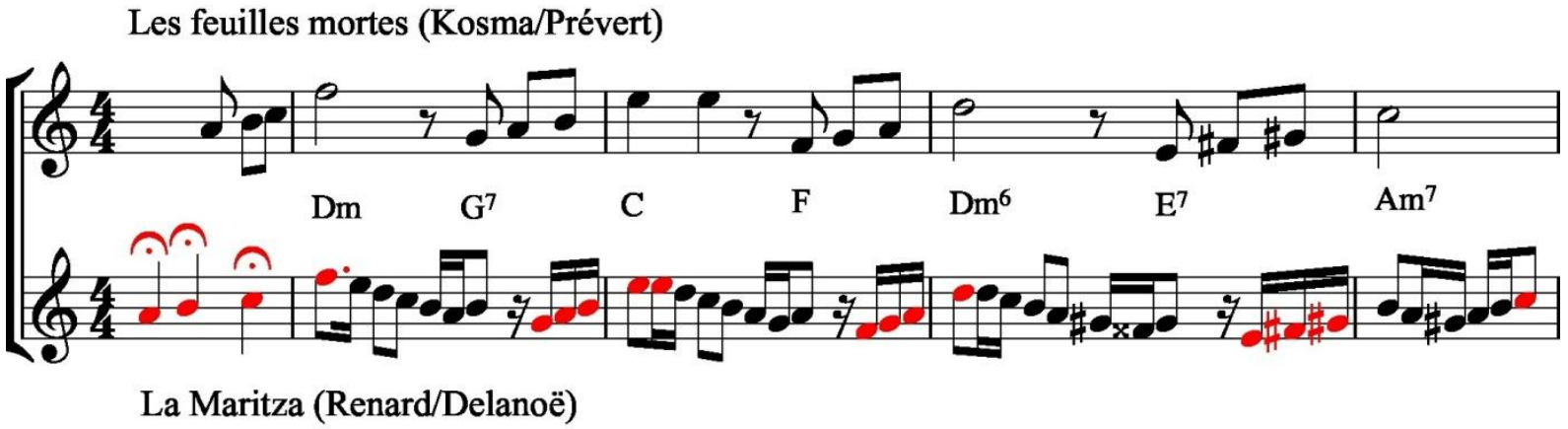
The main part of the existing systems for the comparison of symbolic music are based on string matching algorithms and represent music as sequences of notes.

Identification of Near-Duplicate Music Documents

Musical Plagiarism

In many applications, two strings may not be highly similar in their entirety but may contain regions that are highly similar. In this case, the problem is to find and extract a pair of regions, one from each of the two given strings, that exhibits high similarity. This is called local alignment or local similarity problem. The computation of a local similarity allows us to detect local conserved areas between both sequences. Experiments show that considering local alignment improves the quality of symbolic melodic similarity systems.

Musical Plagiarism

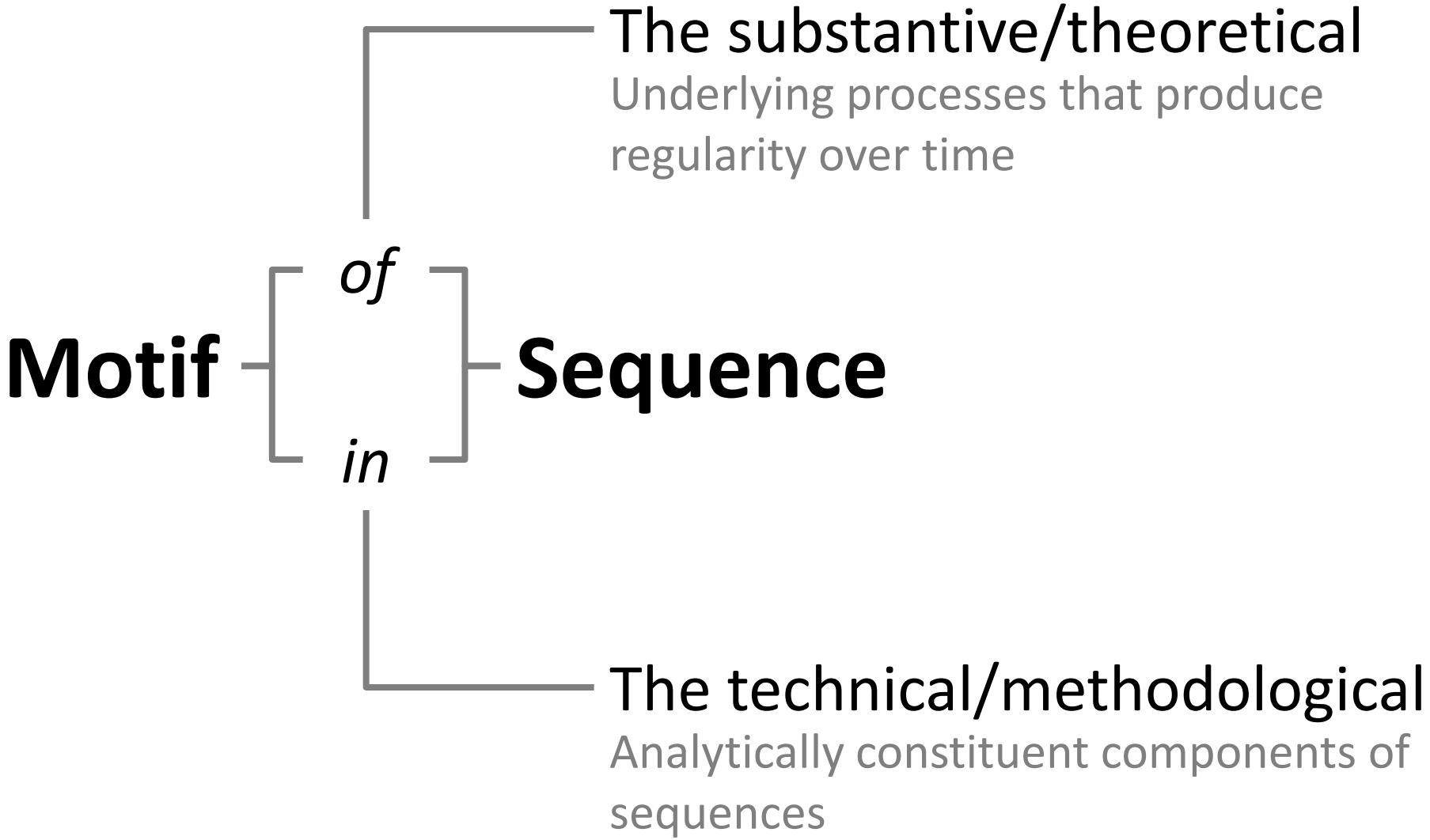


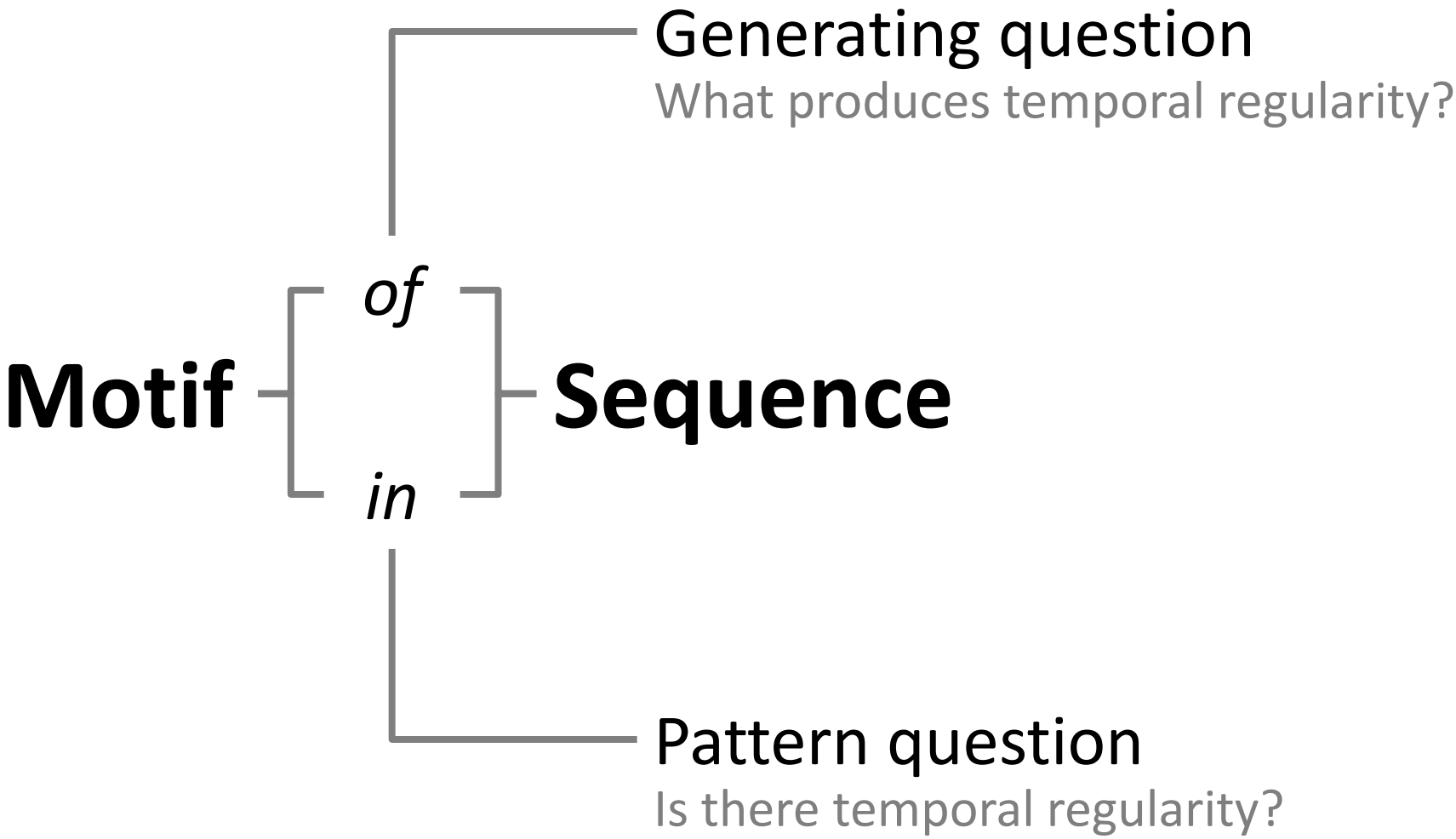
He considered that the chord progression is the same for the refrains of the two musical pieces and that all the notes inserted in *La Maritza* could be considered as ornaments (musical flourishes that are not necessary to the overall melodic or harmonic line). Thus, even if few notes are common to the two musical piece, they are important regarding the harmony.

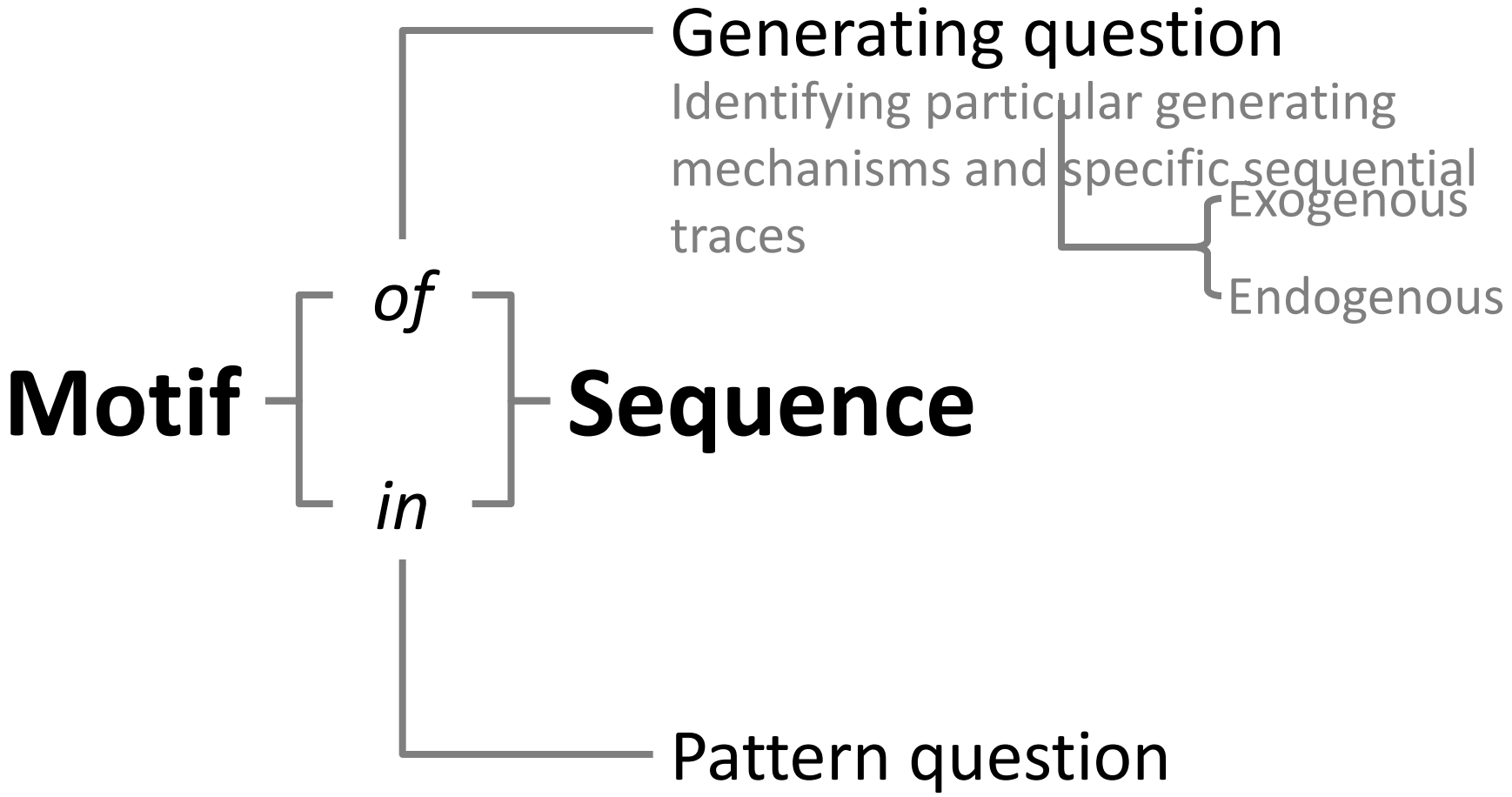
Musical Plagiarism

Templates Repurposed

Motif — The possibility of common subsequences


The substantive/theoretical
Underlying processes that produce regularity over time
Motif
of
in
Sequence
The technical/methodological
Analytically constituent components of sequences


Generating question
What produces temporal regularity?
Motif
of
in
Sequence
Pattern question
Is there temporal regularity?


Generating question
Identifying particular generating mechanisms and specific sequential traces
Exogenous
Endogenous
of
Motif
Sequence
in
Pattern question

Questions

- What are the fundamental building blocks?
- How do they combine to form larger structures?
- Do networks which share the same building blocks also share the same combinations of these blocks?

2 Techniques

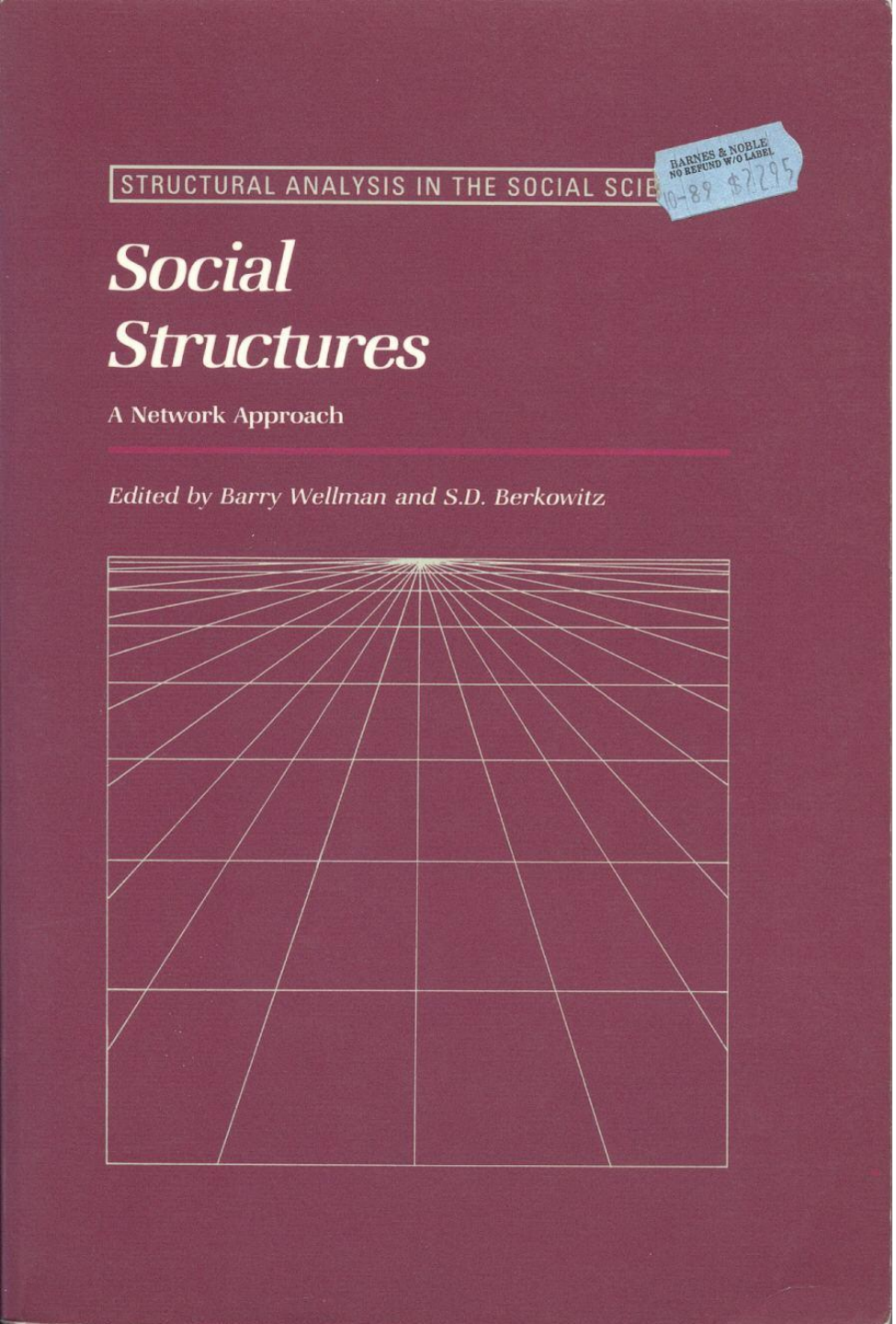
- Two basic techniques have been proposed for identifying network motifs. They attempt to determine the significance of all or many subgraphs of a given size by comparing their frequency in a given network to their frequency in a random ensemble of networks with similar properties to the original.
  - exact counting (complete enumeration)
  - subgraph (subsequence) sampling

2 Techniques

- To determine which subgraphs are motifs, subgraph sampling is an effective and efficient approach, and has been used to evaluate the significance of larger subgraphs than can be evaluated by the exact counting method.

Cut and Sew

- How to cut
  - Complete(d) sequence with beginning, middle, and ending?
  - Boundary? – width
  - Granularity? – arbitrary/malleable grid
- How to sew
  - Script?
  - Signature?



Lessons Borrowed

- Meta-theoretical parallels
- Practical similarities

Structural Analysis: From Method and Metaphor To Theory and Substance

Barry Wellman

(1988)

From Method and Metaphor

To Theory and Substance

A mixed theory of terms and substances

A paradigm

Lessons Borrowed

"[It] does not derive its power from the partial application of this concept or that measure. It is a comprehensive paradigmatic way of taking social structure seriously by studying directly how patterns of ties allocate resources in a social system. Thus, its strength lies in its integrated application of theoretical concepts, ways of collecting and analyzing data, and a growing, cumulating body of substantive findings."

- To pose new intellectual questions
- To collect new types of evidence
- To provide new ways to describe and analyze social structures