UNE NOUVELLE APPROCHE MIXTE D'ENRICHISSEMENT DE DIMENSIONS DANS UN SCHÉMA MULTIDIMENSIONNEL EN CONSTELLATION

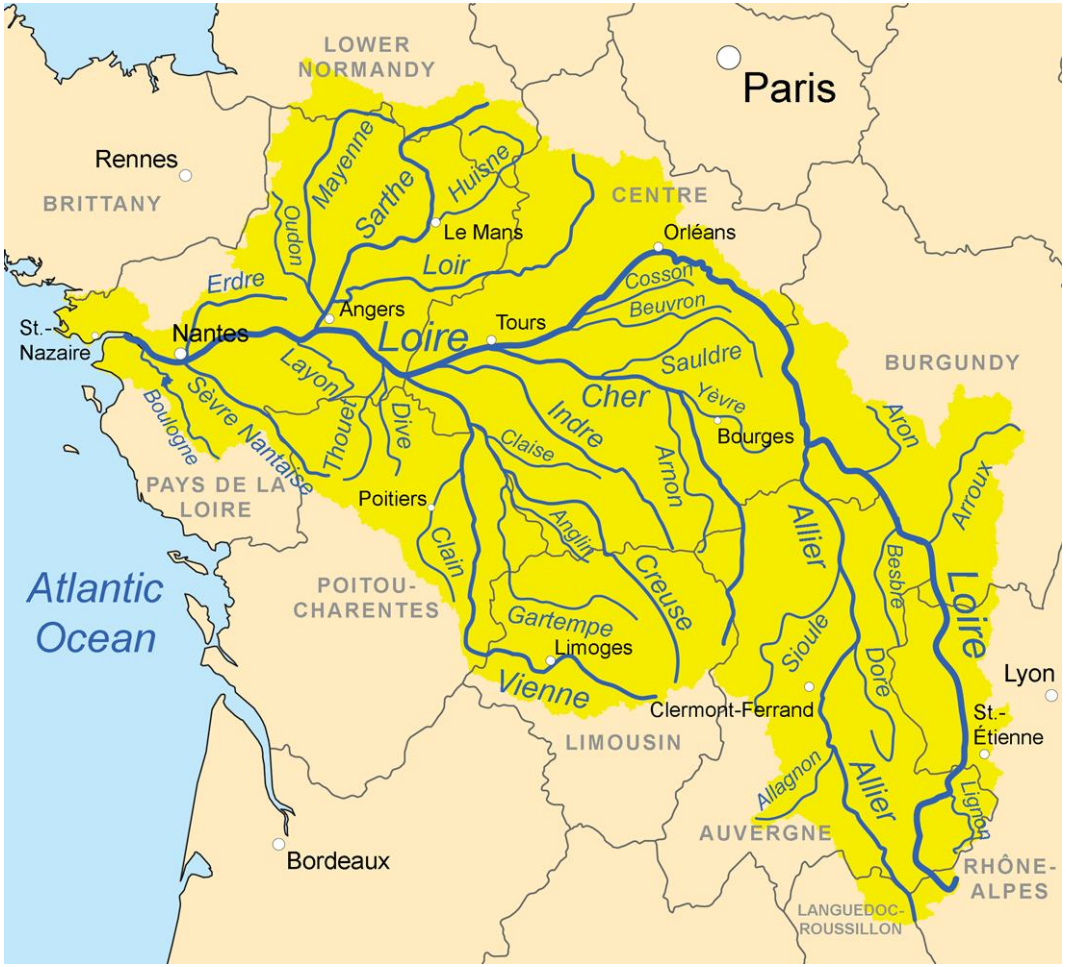
Lucile Sautot
Sandro Bimonte
Ludovic Journaux
Bruno Faivre

Plan

- Introduction et problématique
- Enrichissement de dimensions avec des données factuelles
- Construction automatique de hiérarchies
- Conclusion et perspectives

Introduction

Studying Bird Communities Along a River

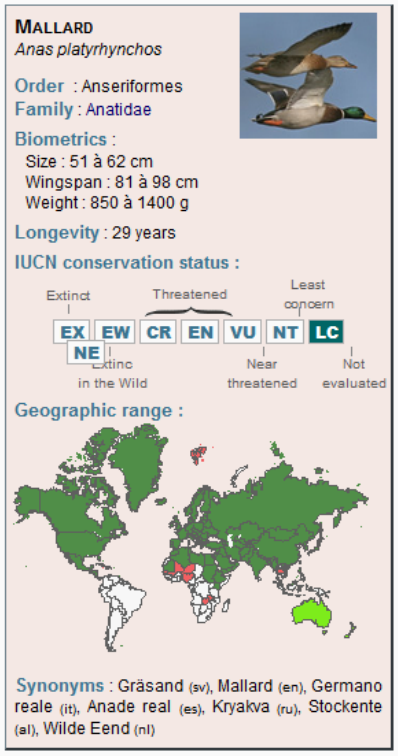


Studying Bird Communities Along a River

Census Point

2.5

Bird Species



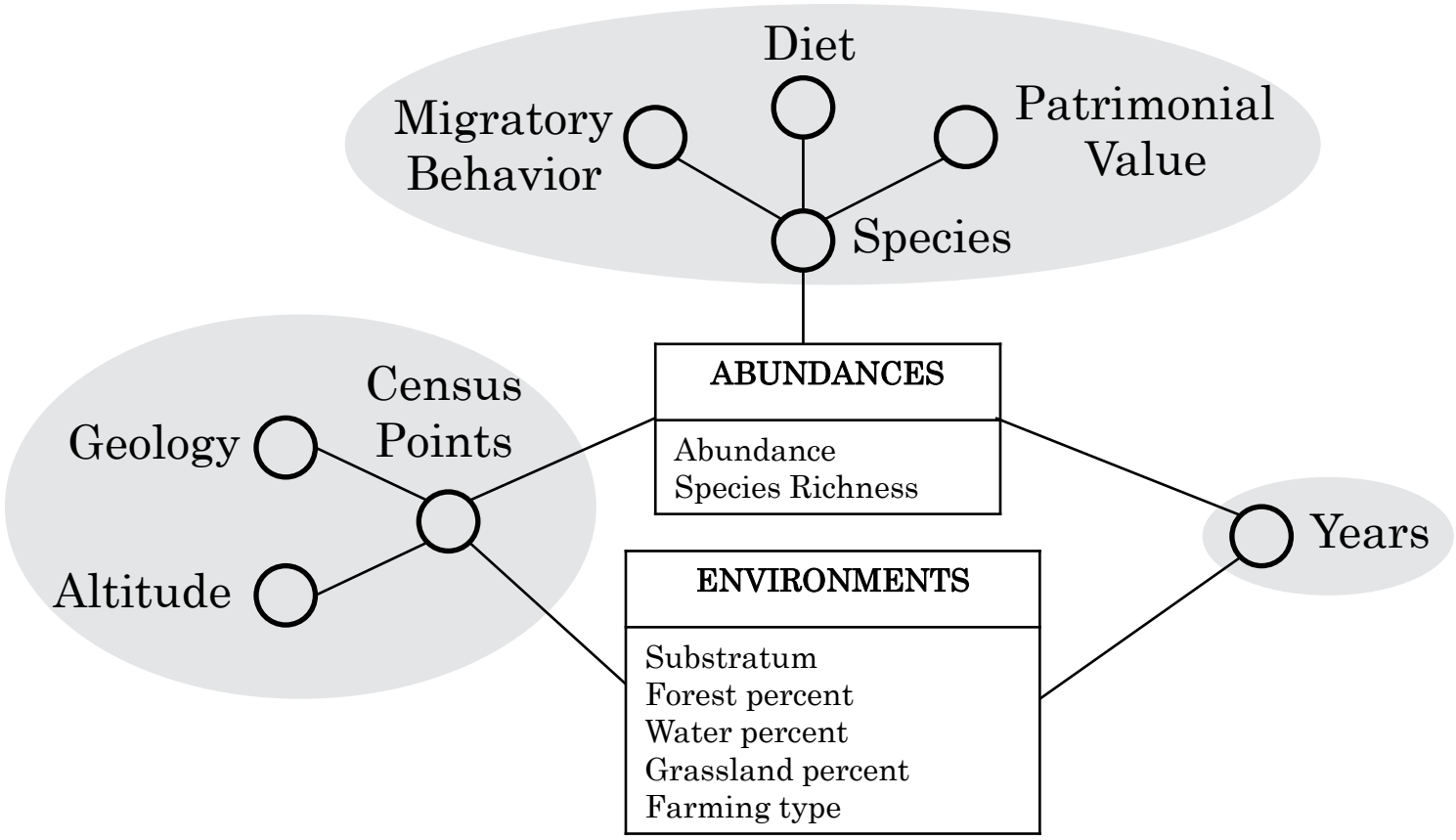
Years

1990

Problematic

Can we create automatically
an OLAP system,
which allows
ecological data analysis ?

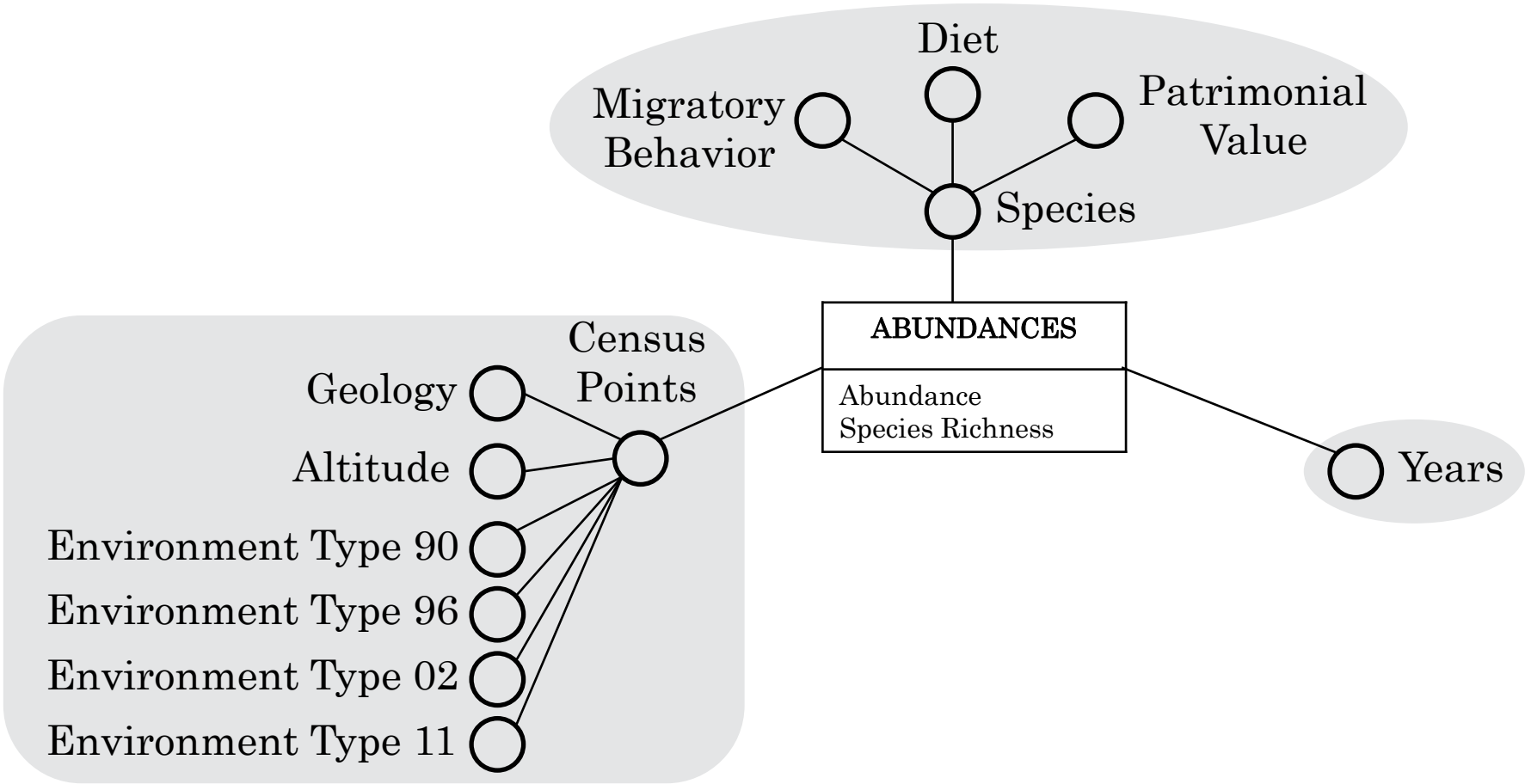
Automatic Modeling



Romero, O. & Abello, A.
Automatic validation of requirements to support multidimensional design
Data and Knowledge Engineering, 2010, 69, 917-942

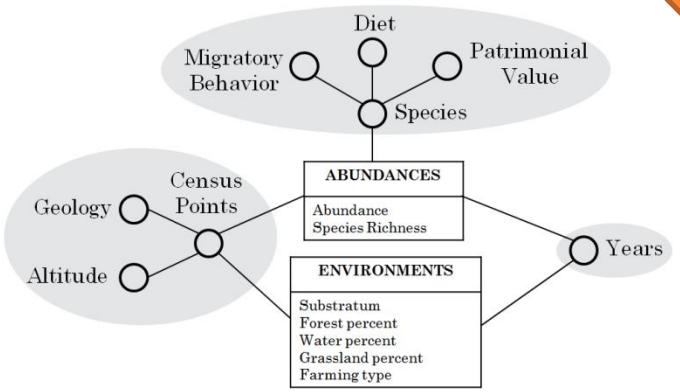
Usman, M.; Pears, R. & Fong, A.
A data mining approach to knowledge discovery from multidimensional cube structures
Knowledge-Based Systems, 2013, 40, 36-49

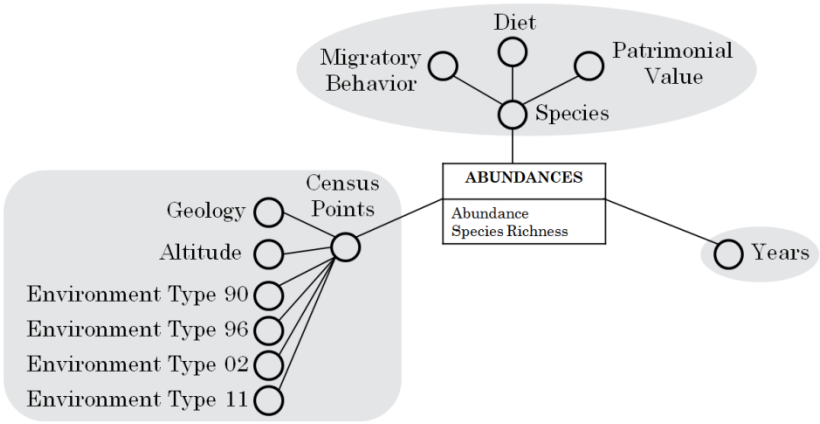
Modeling Complex Data



Miquel, M.; Bédard, Y.; Brisebois, A.; Pouliot, J.; Marchand, P. & Brodeur, J.
Modeling Multi-dimensional Spatio-Temporal Data Warehouses in a Context of Evolving Specifications
International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences, NATURAL RESOURCES CANADA, 2002, 34, 142-147

Our Proposal





Literature Review

- **Messaoud et al., 2004** : agregating facts with the AHC.
- **Favre et al., 2006** : users define rules to build new hierarchies
- **Bentayeb, 2008** : creating new levels in a hierarchy with K-means.
- **Hubert and Teste, 2009** : an operator transforming hierarchies during an OLAP querying session.
- **Leonhardi et al., 2010** : users defines new dimensions during an OLAP querying session.
- **Ceci et al., 2011** : using AHC for integration of continuous data into a dimension.

Dimension Enrichment with Factual Data

PRELIMINARIES

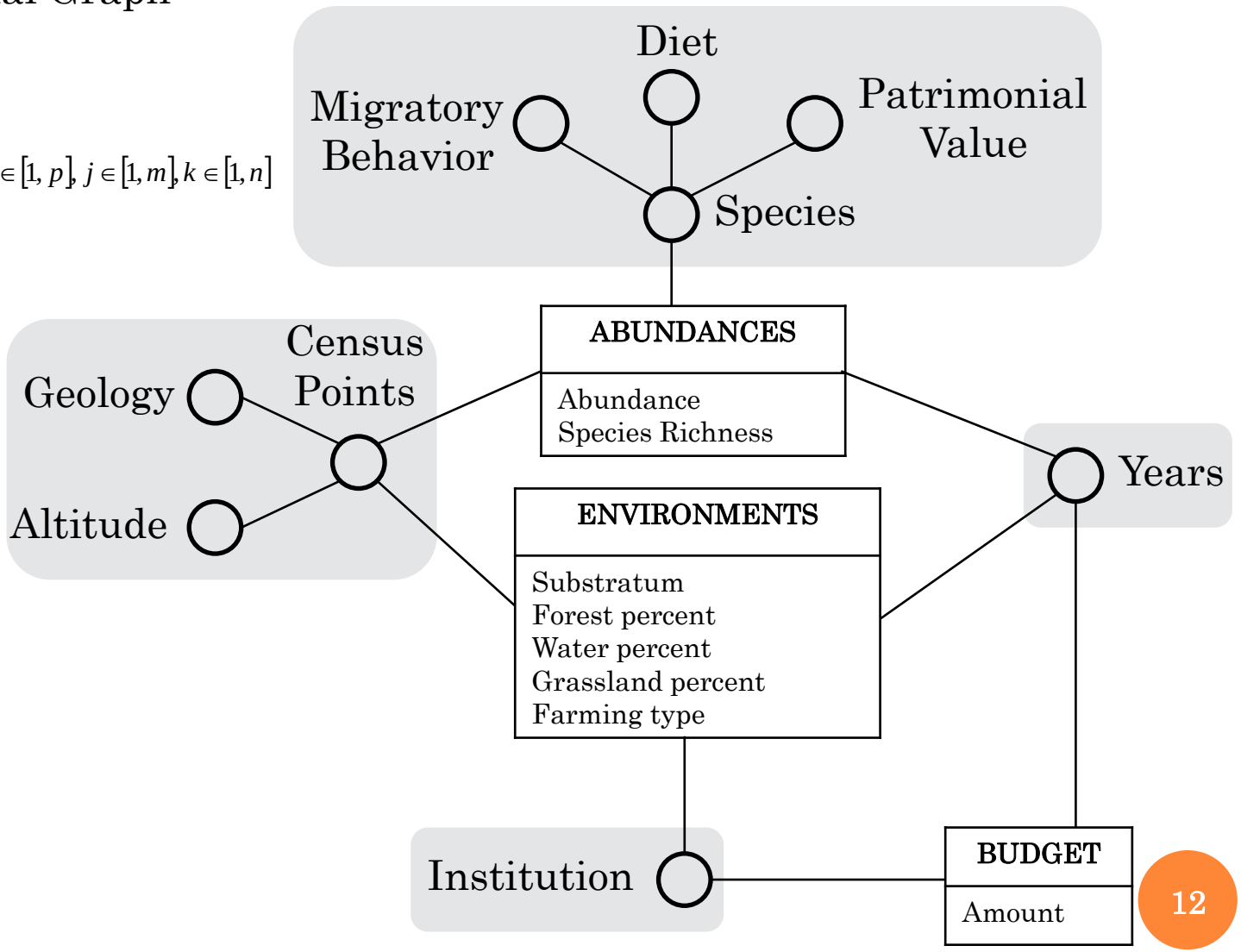
Multidimensional Graph:

$M_G = \langle D, F, A \rangle$

$D = \{d_1, \ldots, d_n\}$

$F = \{f_1, \ldots, f_m\}$

$A = \{a_1, \ldots, a_p \mid a_i = (f_j, d_k), i \in [1, p], j \in [1, m], k \in [1, n]\}$



PRELIMINARIES

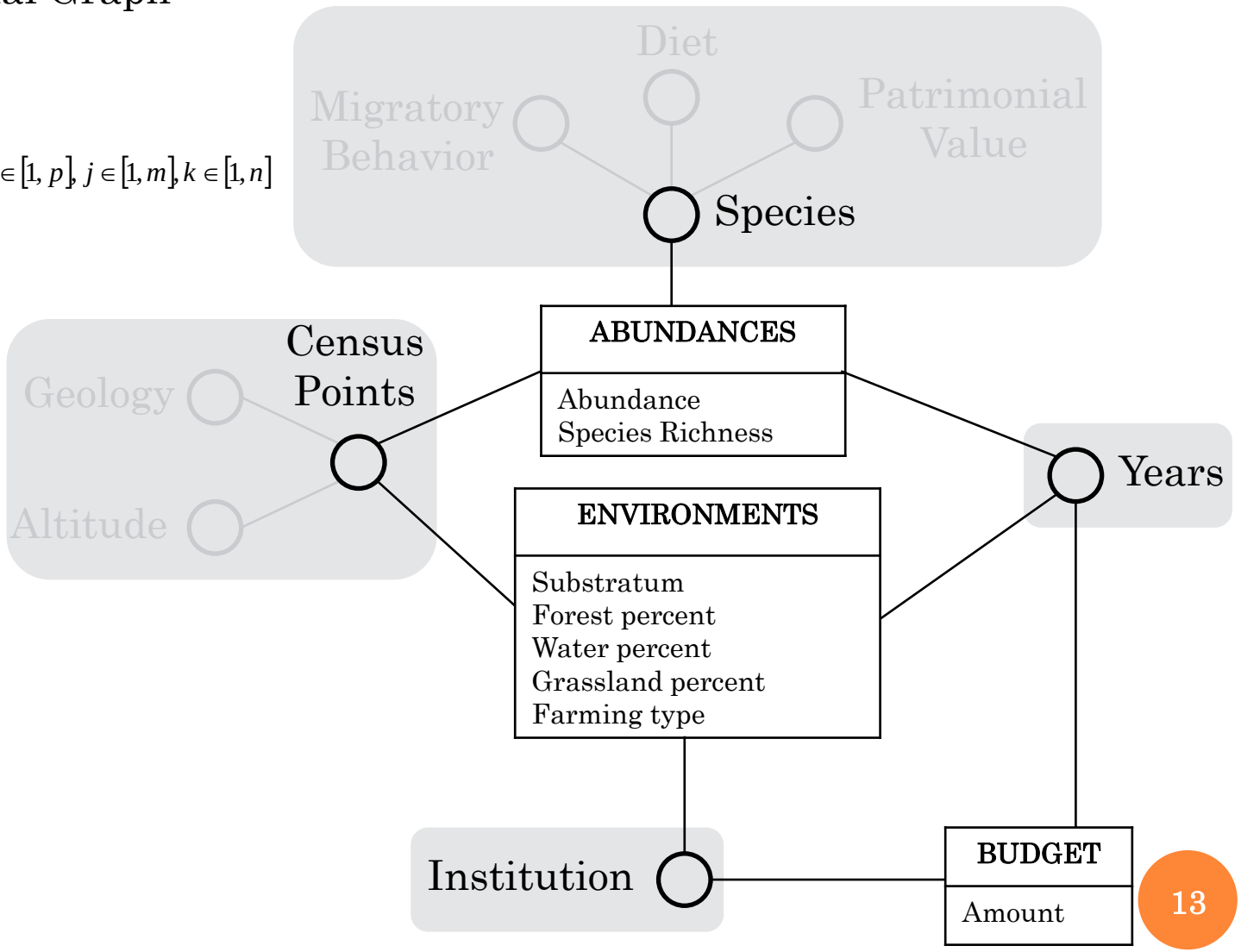
Multidimensional Graph:

$M_G = \langle D, F, A \rangle$

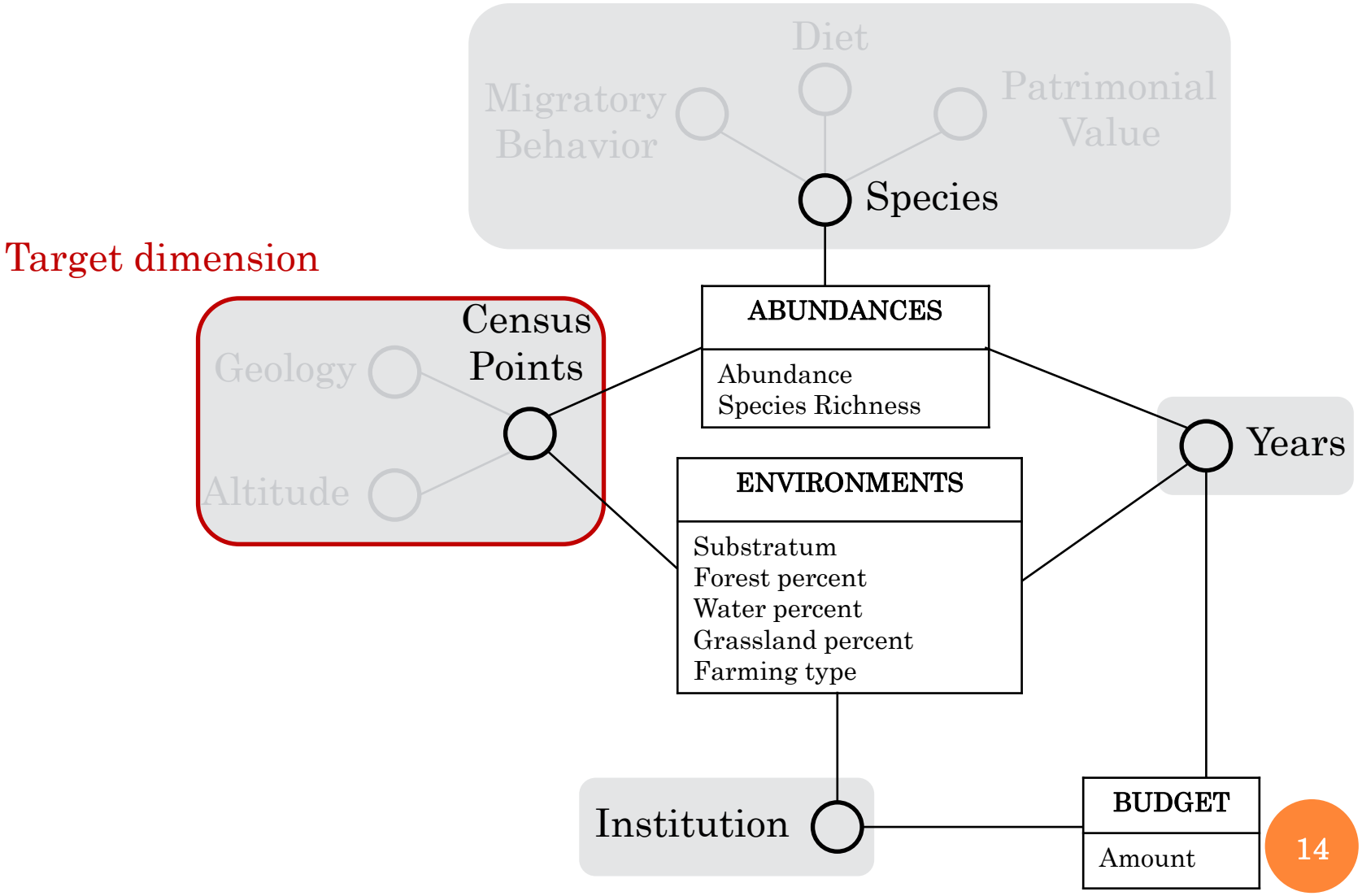
$D = \{d_1, \ldots, d_n\}$

$F = \{f_1, \ldots, f_m\}$

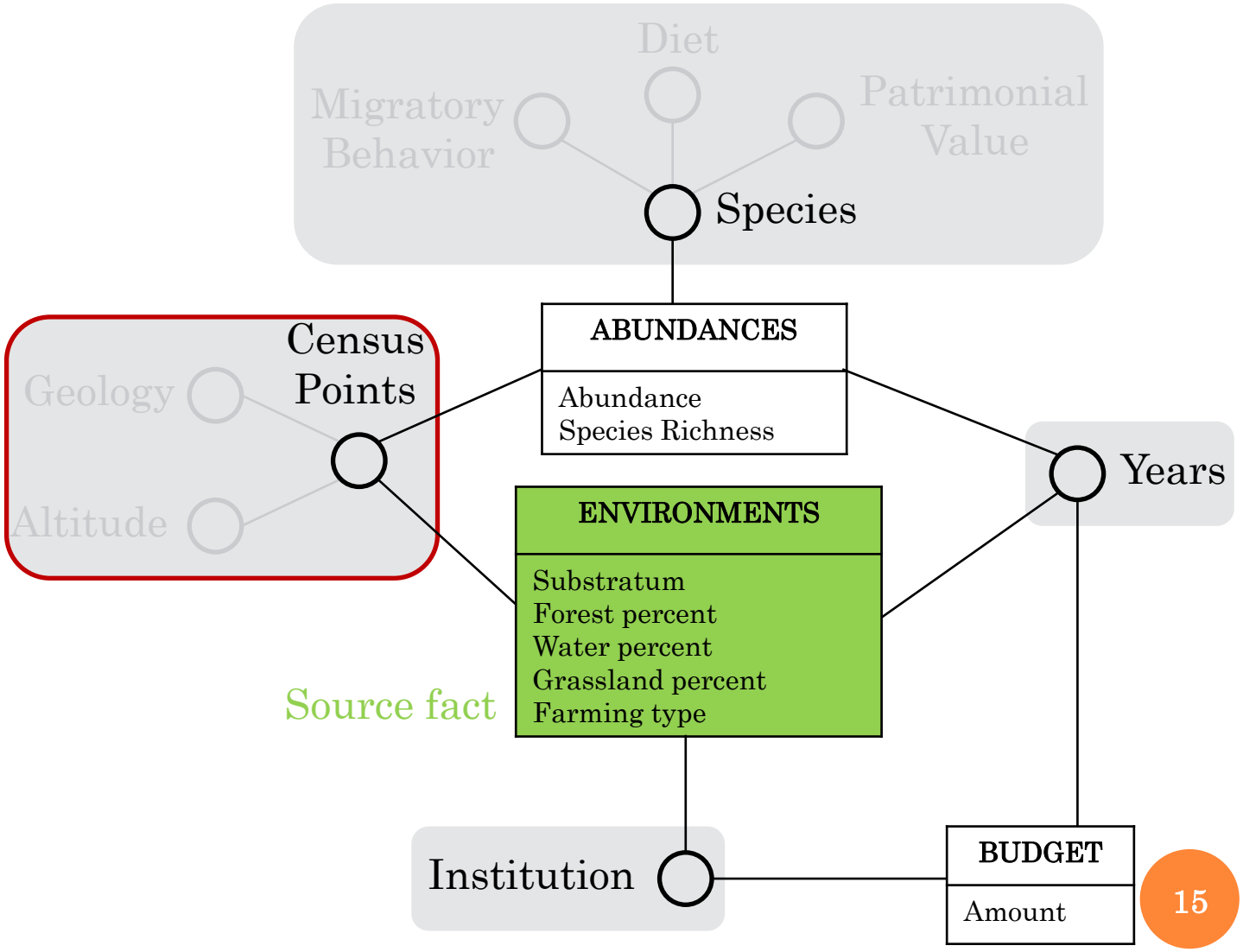
$A = \{a_1, \ldots, a_p \mid a_i = (f_j, d_k), i \in [1, p], j \in [1, m], k \in [1, n]\}$



Inputs


Diet
Migratory Behavior
Patrimonial Value
Species
Target dimension
Census Points
Geology
Altitude
ABUNDANCES
Abundance
Species Richness
Years
ENVIRONMENTS
Substratum
Forest percent
Water percent
Grassland percent
Farming type
Institution
BUDGET
Amount

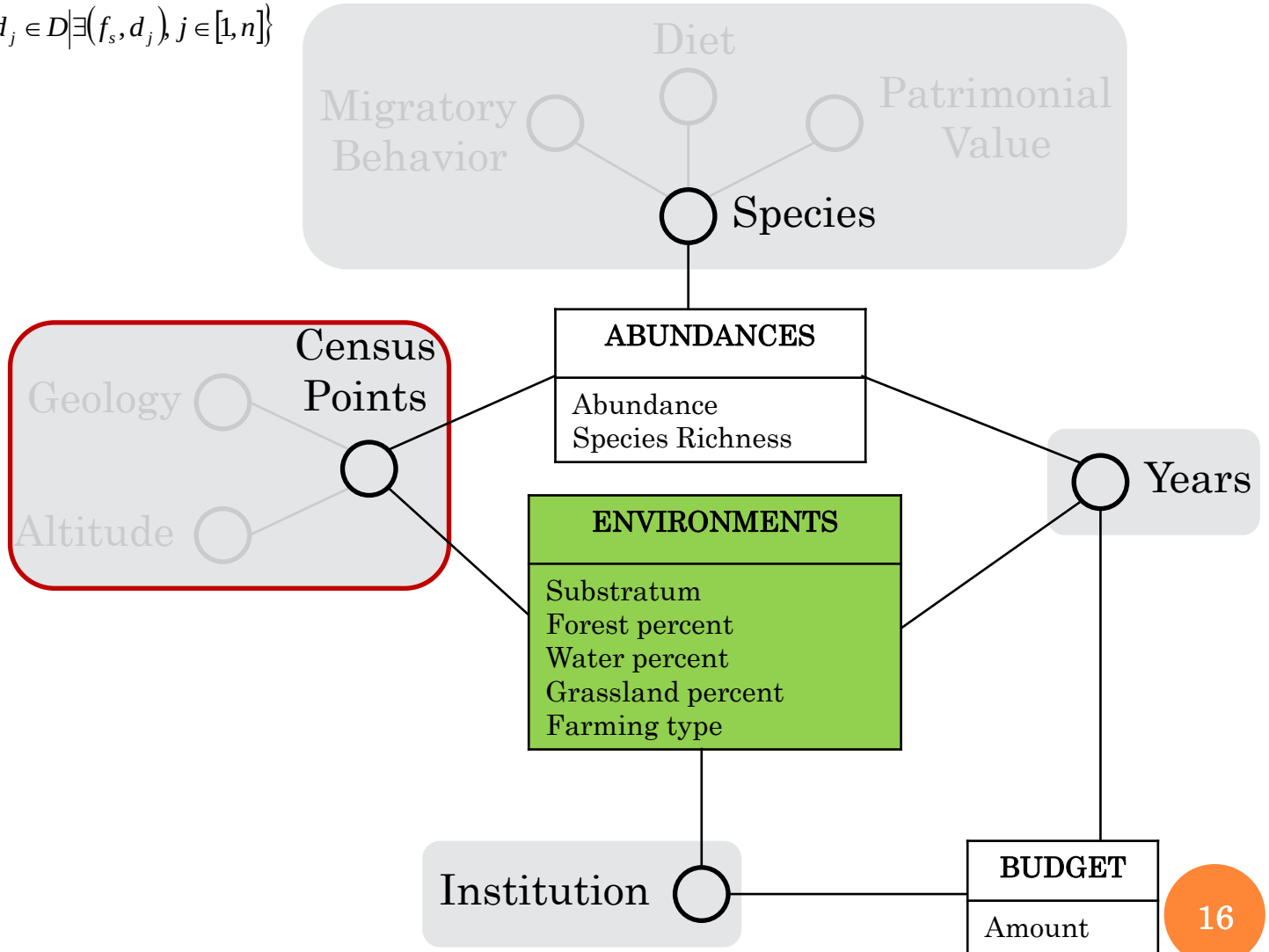
Inputs


Diet
Migratory Behavior
Patrimonial Value
Species
ABUNDANCES
Abundance
Species Richness
Census Points
Geology
Altitude
Years
ENVIRONMENTS
Substratum
Forest percent
Water percent
Grassland percent
Farming type
Source fact
Institution
BUDGET
Amount

Selection

Selected items:

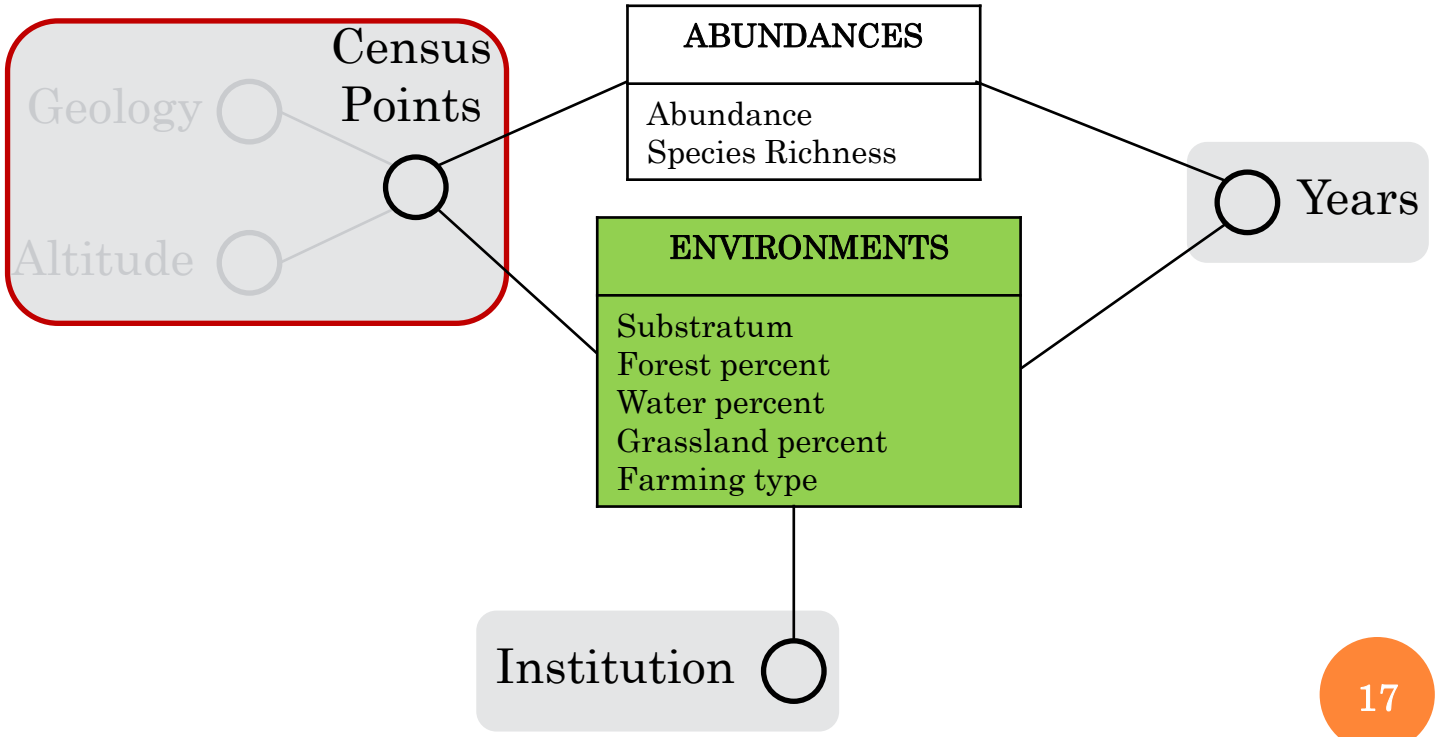
$$\{f_i \in F | \exists (f_i, d_t), i \in [1,m]\} \cup \{d_j \in D | \exists (f_s, d_j), j \in [1,n]\}$$



Selection

Selected items:

$$\{f_i \in F | \exists (f_i, d_t), i \in [1,m]\} \cup \{d_j \in D | \exists (f_s, d_j), j \in [1,n]\}$$

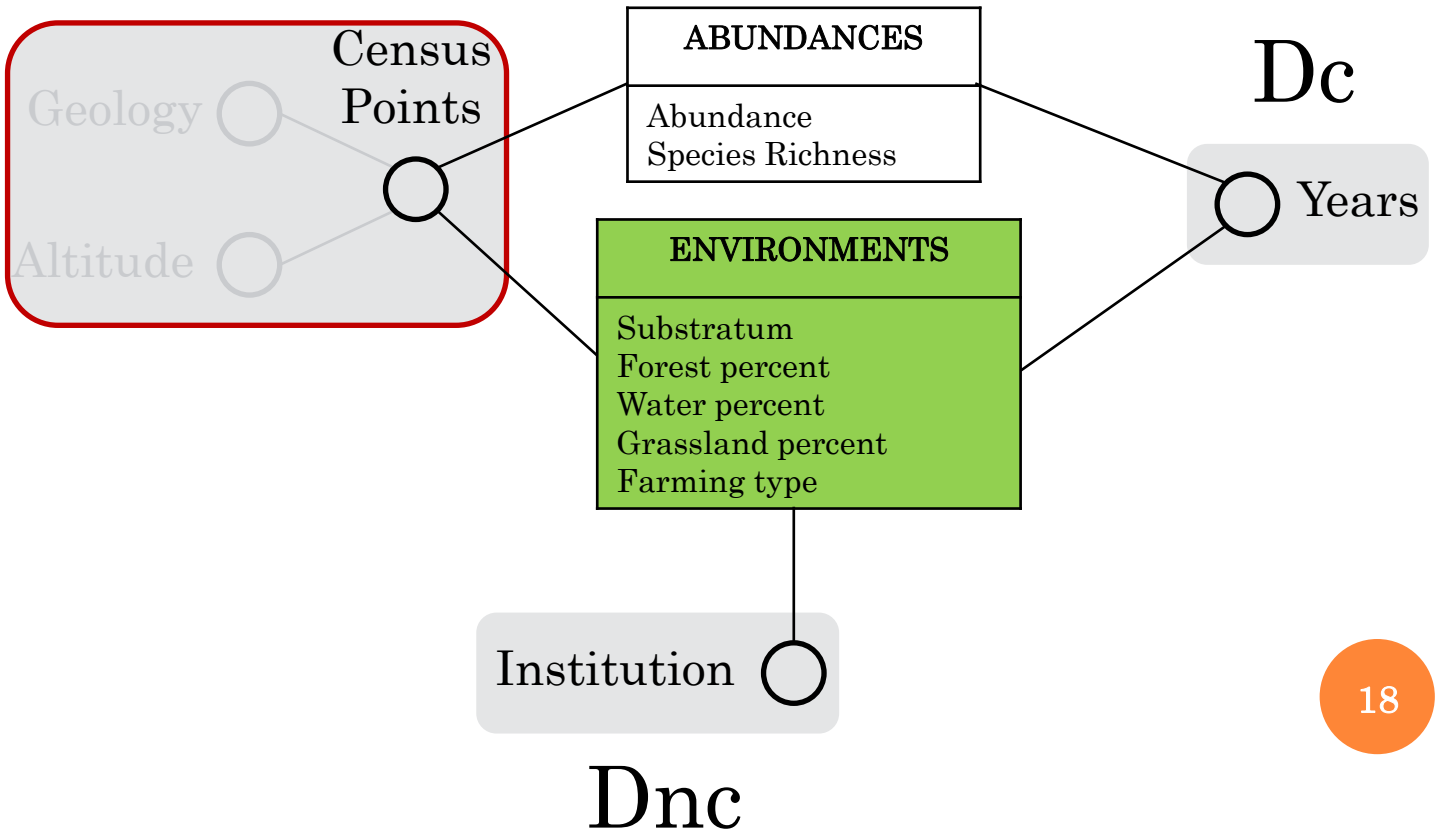


Characterization of Dimensions

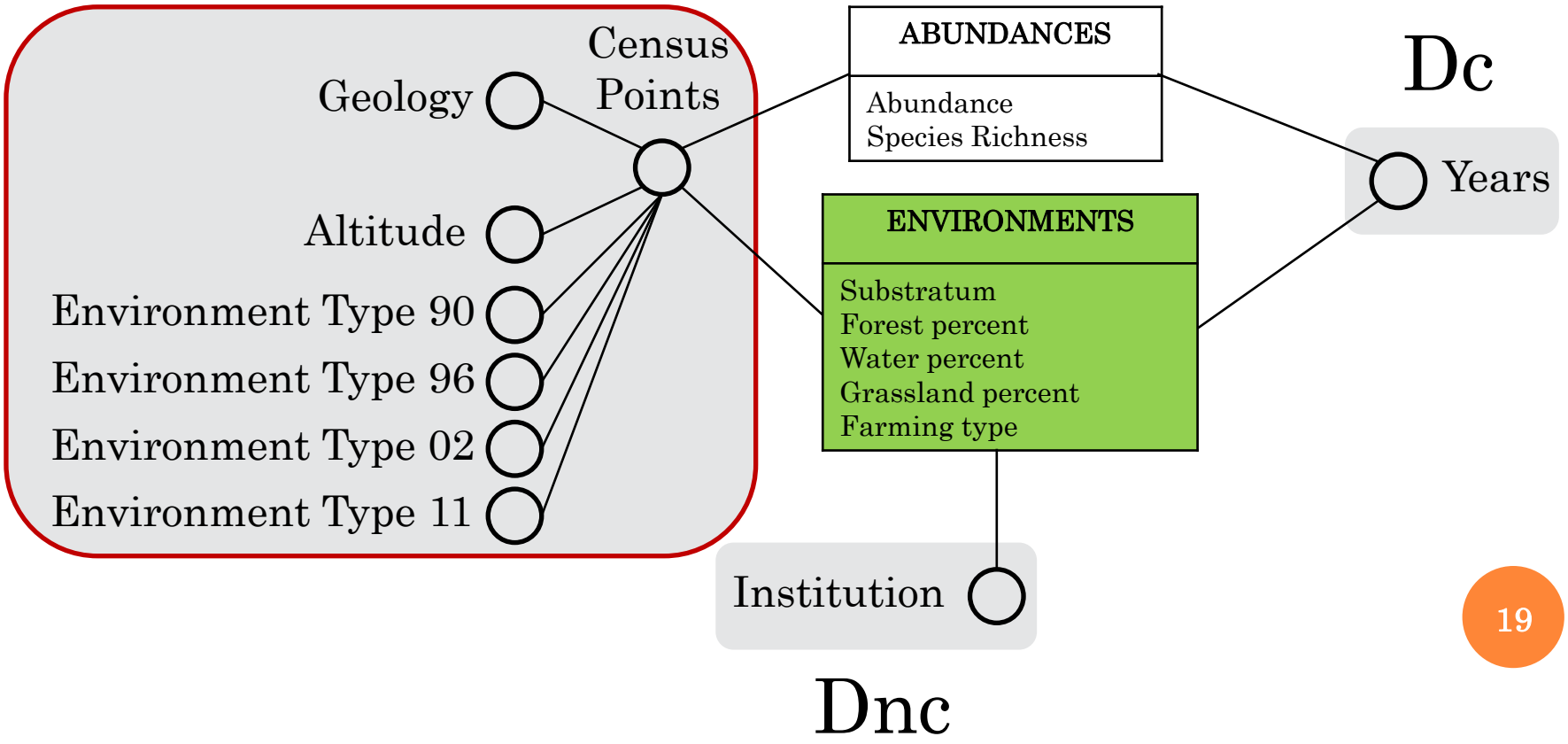
Dimensional sets:

$$D_{nc} = \left\{ d_u^{nc} \middle| \exists! a_u = \left(d_u^{nc}, f \right) \right\}$$

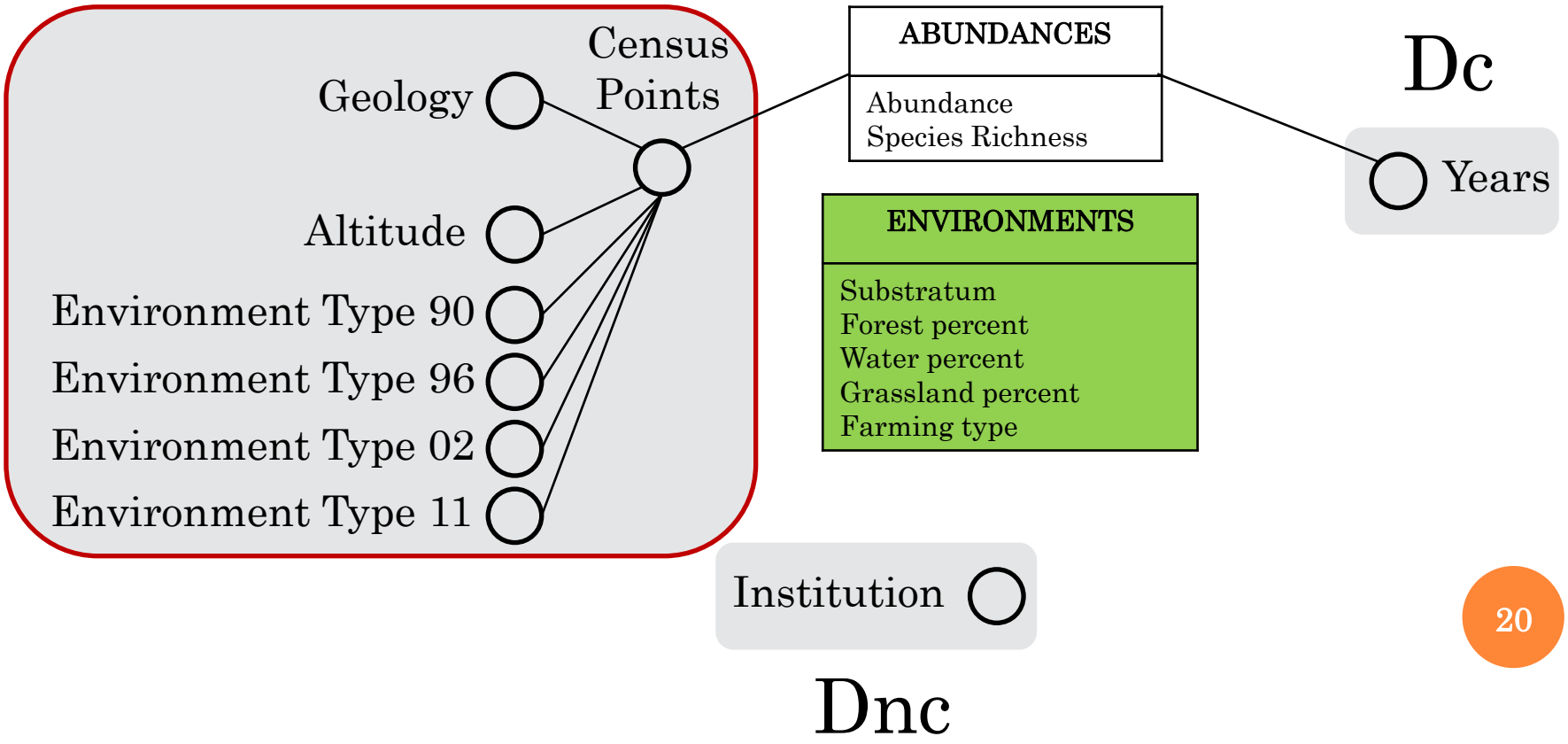
$$D_c = D' - \left(D_{nc} \cup \{ d_t \} \right)$$



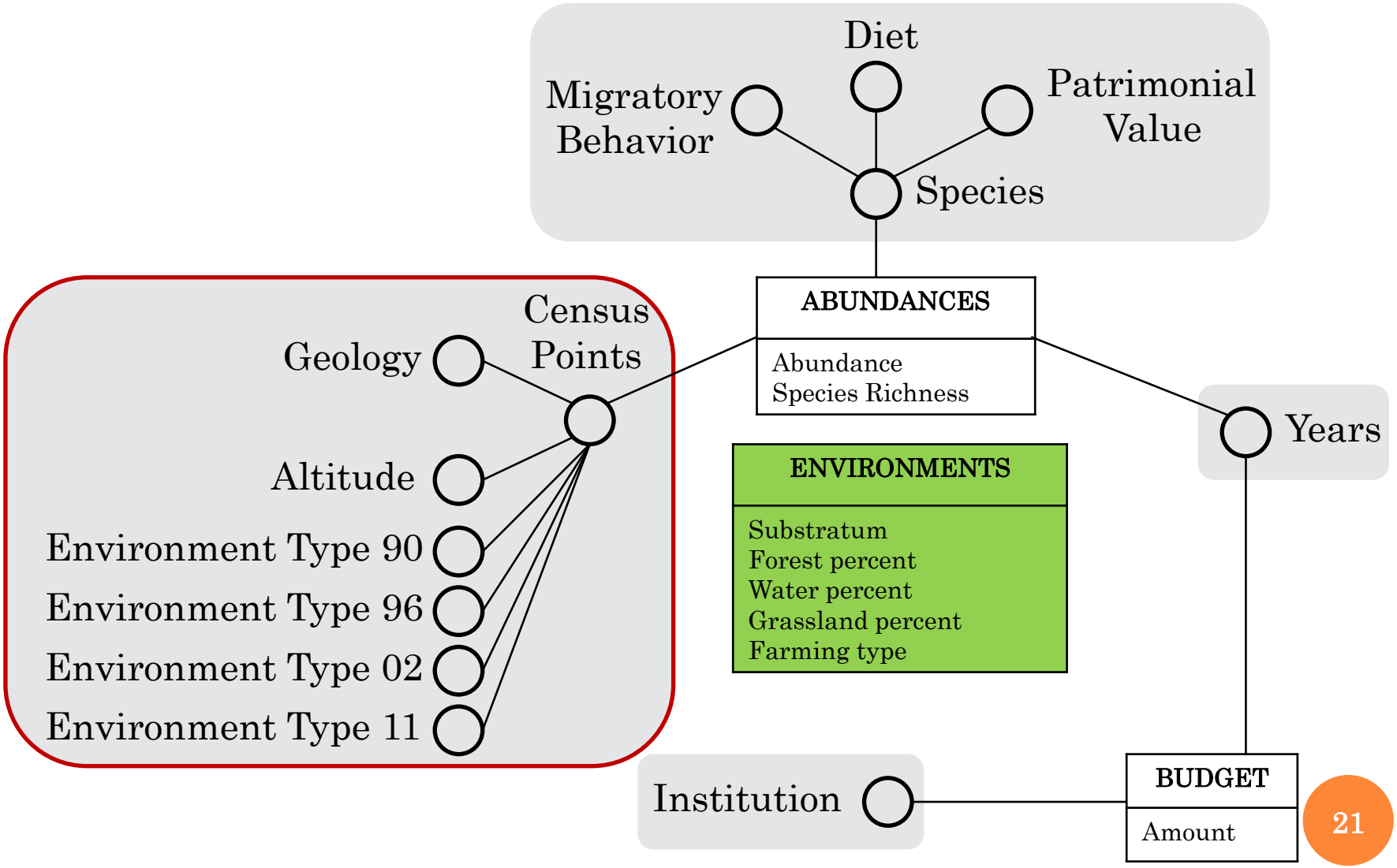
Hierarchy Calculation



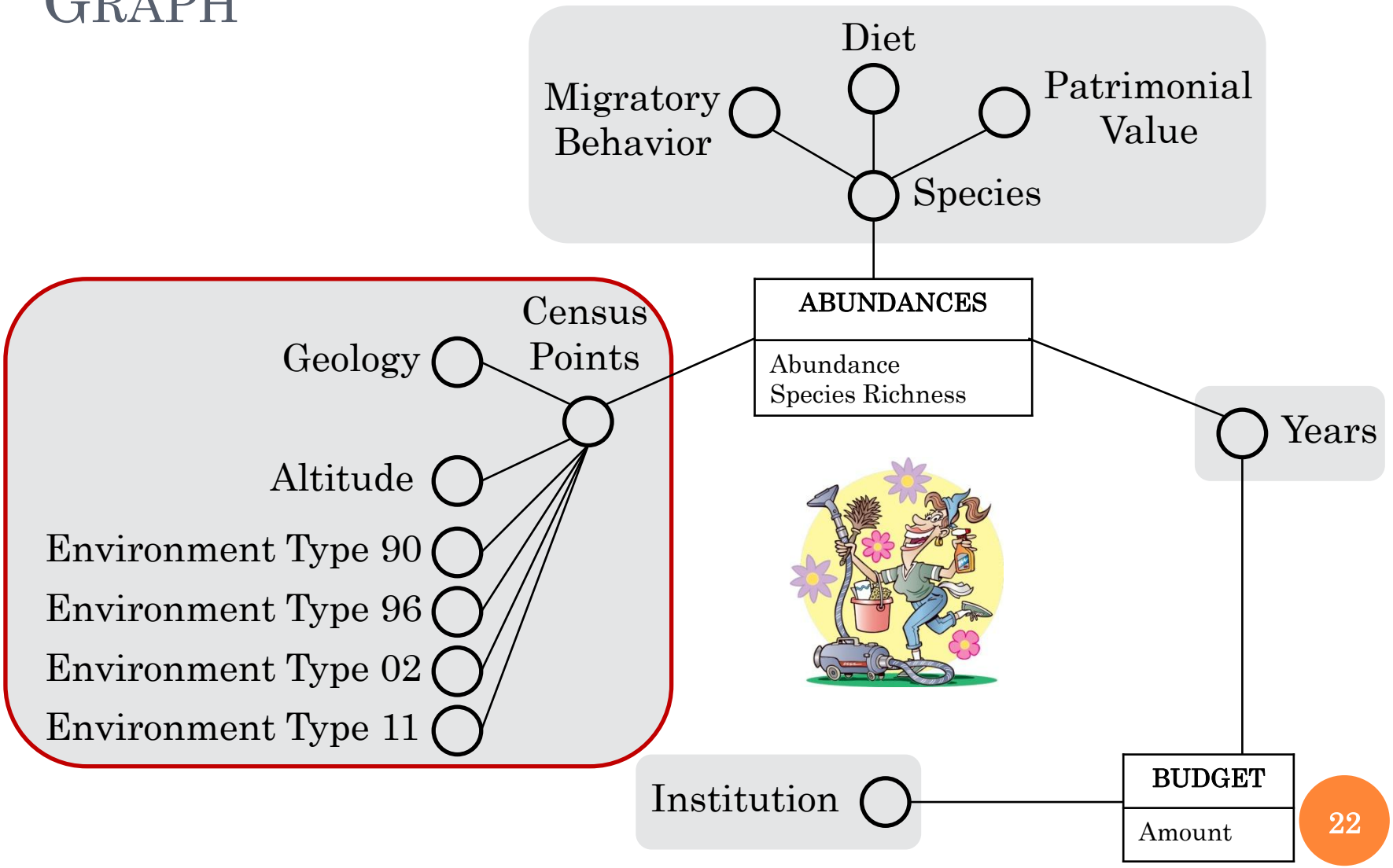
Deletion of Arcs from the Source Fact



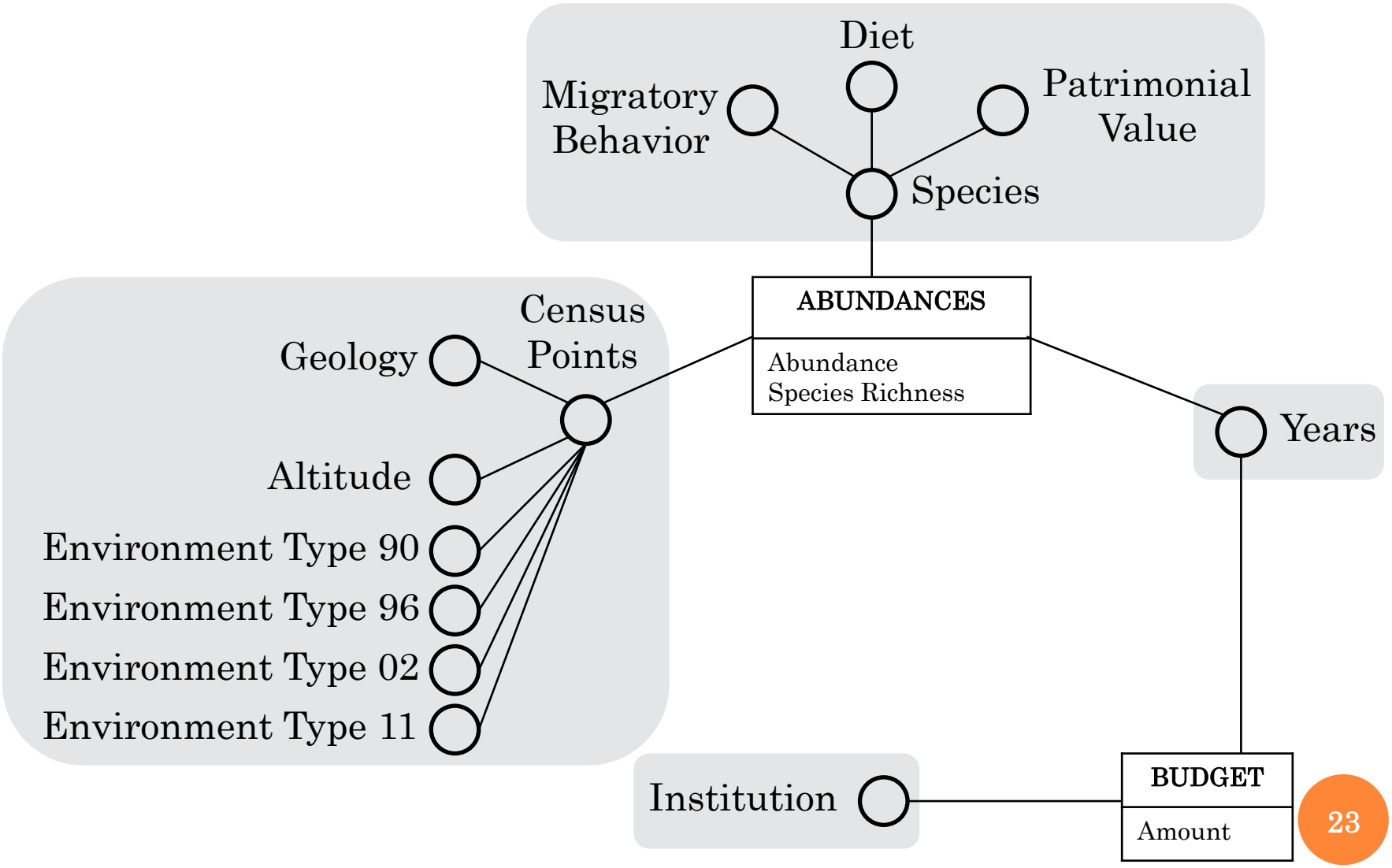
Come Back to the Initial Graph



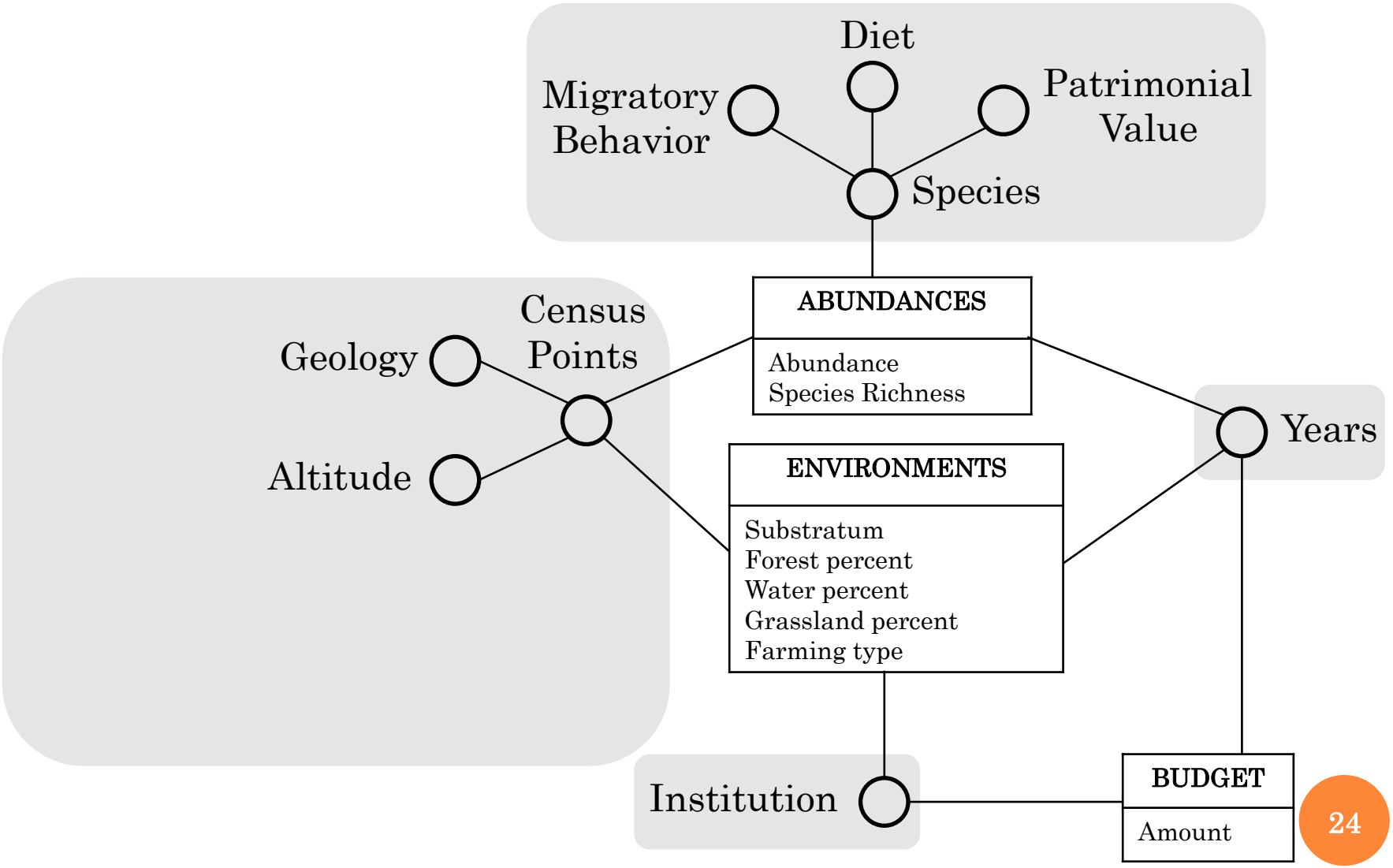
Cleaning the Multidimensional Graph



Refined Result

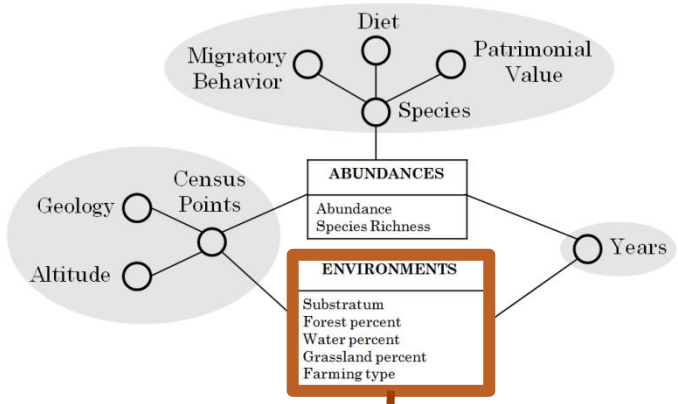


Initial Graph



Automatic Building of Hierarchies

The Source facts



Year	Census Point	Water percent	Forest percent	Farming type
1990	1	15.2	45	A
1990	2	15.3	56	A
1990	3	15	50	A
1990	4	12	21	B
2002	1	15.2		A
2002	2	15		A
2002	3	15.8		A
2002	4	16		A

Instances of a Context

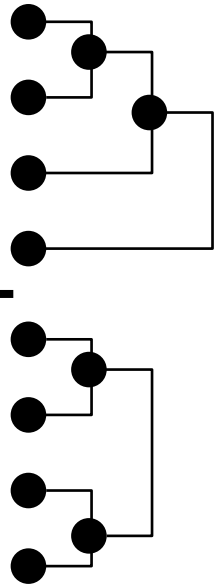
Year	Census Point	Water percent	Forest percent	Farming type
1990	1	15.2	45	A
1990	2	15.3	56	A
1990	3	15	50	A
1990	4	12	21	B
2002	1	15.2		A
2002	2	15		A
2002	3	15.8		A
2002	4	16		A

Instances of Context

Classified members

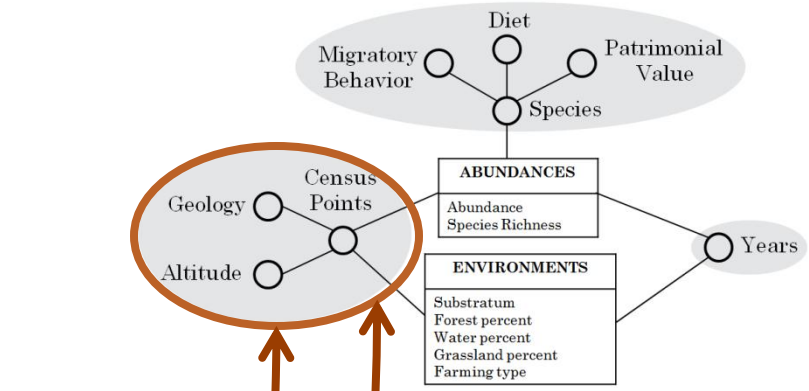
Hierarchical Clustering

Year	Census Point	Water percent	Forest percent	Farming type
1990	1	15.2	45	A
1990	2	15.3	56	A
1990	3	15	50	A
1990	4	12	21	B
2002	1	15.2		A
2002	2	15		A
2002	3	15.8		A
2002	4	16		A



Hierarchy Integration

Year	Census Point	Water percent	Forest percent	Farming type
1990	1	15.2	45	A
1990	2	15.3	56	A
1990	3	15	50	A
1990	4	12	21	B
2002	1	15.2		A
2002	2	15		A
2002	3	15.8		A
2002	4	16		A



Conclusion

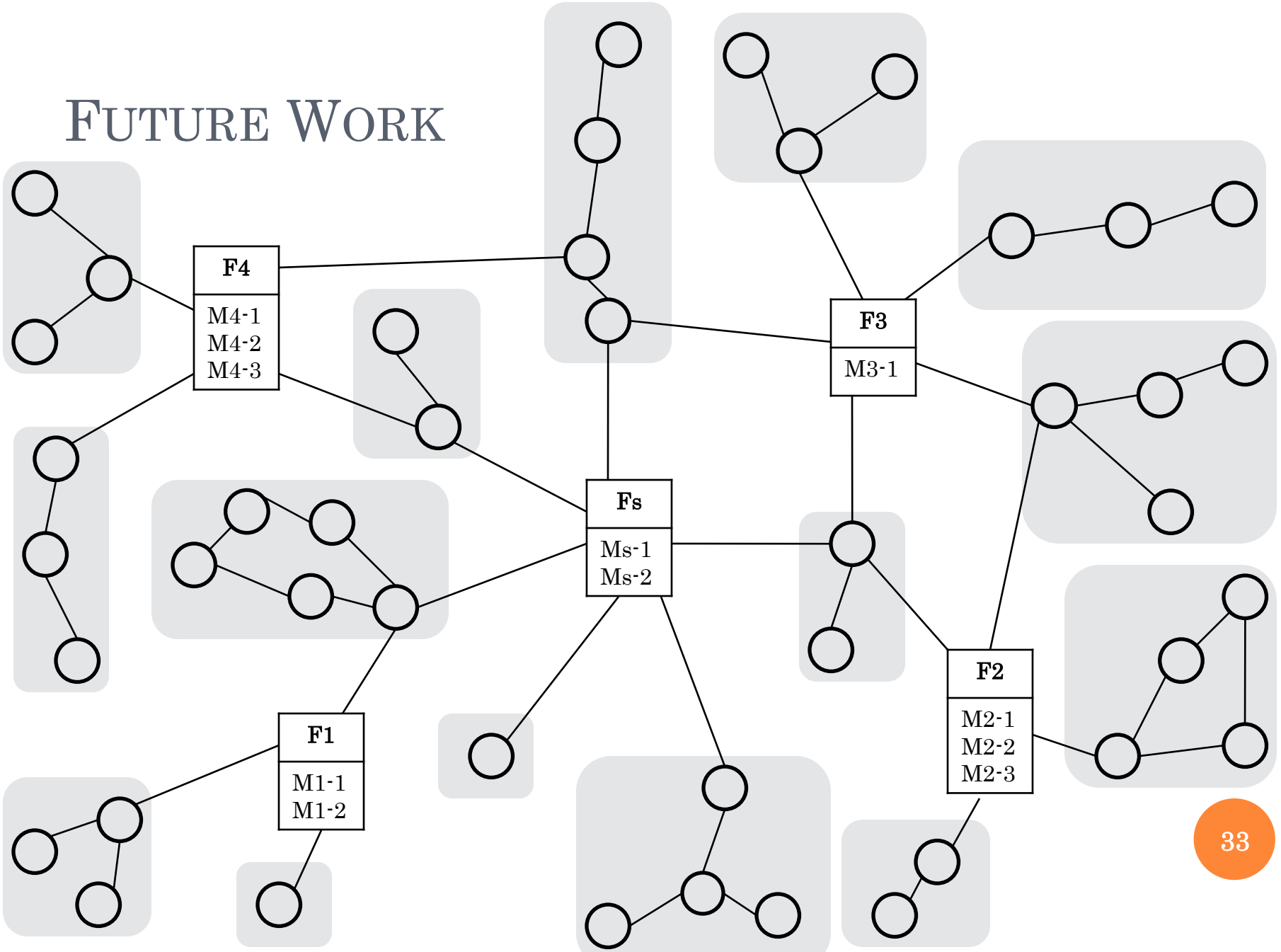
Conclusion

- An algorithm enriching a dimension with factual data
- Taking account of the context is necessary: Contextual Data Warehouse
- Automatic design of Data Warehouse for complex data

Conclusion

- An algorithm enriching a dimension with factual data
- Taking account of the context is necessary: Contextual Data Warehouse
- Automatic design of Data Warehouse for complex data
- *But only one contextual dimension*

Future Work



Acknowledgments

Financial Support

Thank you for your Attention

Questions ?