CAUSAL REASONING APPLIED TO SENSORY ANALYSIS: THE CASE OF THE ITALIAN WINE

SILVIA GOLIA, EUGENIO BRENTARI AND MAURIZIO CARPITA

UNIVERSITY OF BRESCIA - ITALY

MARCH 23, 2016

AIM OF THE STUDY

IN THIS STUDY THE CAUSAL REASONING IS APPLIED TO THE SENSORY ANALYSIS FIELD IN ORDER TO IDENTIFY THE FACTORS THAT HAVE A DIRECT INFLUENCE IN DETERMINING THE QUALITY OF THE ITALIAN WINE.

THE DATABASE WAS CREATED USING THE DATA PRODUCED BY **ALTROCONSUMO** FOR ITS ANNUAL PUBLICATION **GUIDA VINI**.

A COMPLETE PARTIALLY DIRECTED ACYCLIC GRAPH (CPDAG) IS PROPOSED AS AN EQUIVALENT CLASS TO WHICH THE CAUSAL STRUCTURE UNDERLYING THE AVAILABLE VARIABLES AND DATA BELONGS.

CAUSAL MODELS

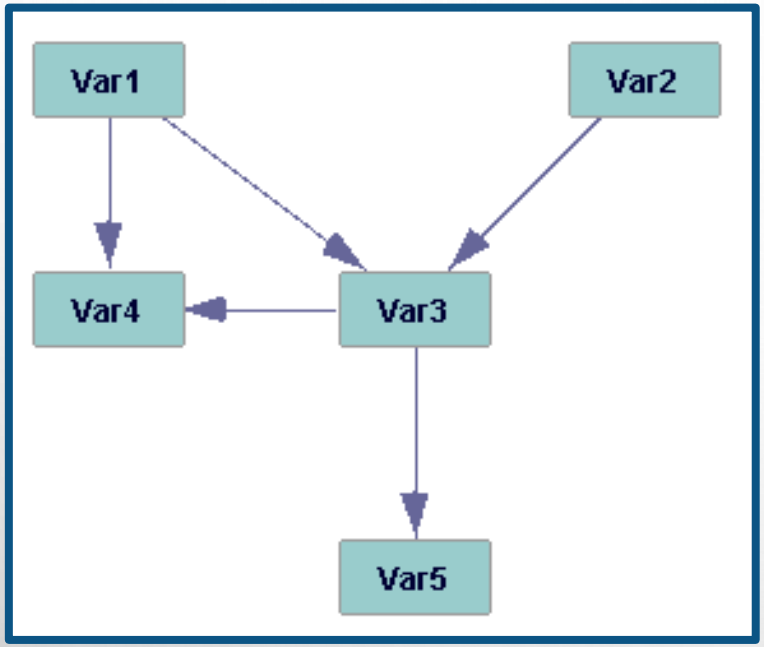
A **CAUSAL MODEL** IS A STATISTICAL MODEL THAT SPECIFIES A SET OF PROBABILITY DENSITIES OVER A GIVEN SET OF VARIABLES AND, FOR EACH MANIPULATION THAT CAN BE PERFORMED ON THE POPULATION, IT ALSO SPECIFIES A SET OF POST-MANIPULATION PROBABILITY DENSITIES OVER THE VARIABLES.

A CAUSAL MODEL IS SPECIFIED IN TWO PARTS: A **STATISTICAL MODEL**, AND A CAUSAL GRAPH THAT DESCRIBES THE CAUSAL RELATIONS BETWEEN VARIABLES (**DIRECTED ACYCLIC GRAPH – DAG**).

THE MOST FREQUENTLY USED CAUSAL MODELS BELONG TO TWO FAMILIES:

- **CAUSAL BAYESIAN NETWORKS**
- **STRUCTURAL EQUATION MODELS**.

CAUSAL MODELS



RECURSIVE SEM

$$\mathbf{Var1} = \alpha_{10} + \varepsilon_1$$

$$\mathbf{Var2} = \alpha_{20} + \varepsilon_2$$

$$\mathbf{Var3} = \alpha_{30} + \alpha_{31}\mathbf{Var1} + \alpha_{31}\mathbf{Var2} + \varepsilon_3$$

$$\mathbf{Var4} = \alpha_{40} + \alpha_{41}\mathbf{Var1} + \alpha_{43}\mathbf{Var3} + \varepsilon_4$$

$$\mathbf{Var5} = \alpha_{50} + \alpha_{53}\mathbf{Var3} + \varepsilon_5$$

BAYESIAN NETWORKS

$$P(\mathbf{Var1}, \mathbf{Var2}, \mathbf{Var3}, \mathbf{Var4}, \mathbf{Var5}) =$$

$$= P(\mathbf{Var1})\, P(\mathbf{Var2})\, P(\mathbf{Var3} \mid \mathbf{Var1}, \mathbf{Var2})\, P(\mathbf{Var4} \mid \mathbf{Var1}, \mathbf{Var3})\, P(\mathbf{Var5} \mid \mathbf{Var3})$$

CAUSAL STRUCTURE

TO FIND THE CAUSAL STRUCTURE REPRESENTED AS A DAG IS A PROBLEM IMPOSSIBLE TO SOLVE WITH OBSERVATIONAL DATA ONLY. NEVERTHELESS, UNDER SUITABLE ASSUMPTIONS SUCH AS CAUSAL SUFFICIENCY, CAUSAL MARKOV CONDITION AND CAUSAL FAITHFULNESS CONDITION, CAUSAL STRUCTURES CAN BE RECOVERED AT LEAST UP TO SOME **EQUIVALENCE CLASS TO WHICH THE TRUE DAG BELONGS**.

ONE CONSIDERS A SET OF VARIABLES ***V*** WHOSE CAUSAL STRUCTURE CAN BE REPRESENTED BY A DAG ***G***.

- **CAUSAL SUFFICIENCY**

  THE SET ***V*** INCLUDES ALL OF THE COMMON CAUSES OF PAIRS IN ***V***, THAT IS THERE ARE NO CONFOUNDERS OF VARIABLES IN ***V*** THAT HAVE BEEN LEFT OUT OF ***V***.

CAUSAL STRUCTURE

- **CAUSAL MARKOV CONDITION**

  EVERY VARIABLE IN ***V*** IS PROBABILISTIC INDEPENDENT OF ITS NON-EFFECTS (NON-DESCENDANTS IN ***G***) CONDITIONAL ON ITS DIRECT CAUSES (PARENTS IN ***G***)

- **CAUSAL FAITHFULNESS (STABILITY) CONDITION**

  NO CONDITIONAL INDEPENDENCE HOLDS UNLESS ENTAILED BY THE CAUSAL MARKOV CONDITION

CAUSAL STRUCTURE

CONSTRAINT-BASED ALGORITHMS

PC-ALGORITHM (SPIRTES ET AL., 2000)

1. USING CONDITIONAL INDEPENDENCE TESTS, AN UNDIRECTED GRAPH IS PROVIDED
2. ORIENT UNSHIELDED COLLIDERS
3. ADD IMPLIED ORIENTATIONS USING MEEK RULES

CONSERVATIVE PC-ALGORITHM (RAMSEY ET AL., 2006)

1. IT MODIFIES STEP 2 IN PC ALGORITHM
2. IT IS ABLE TO GIVE A MORE CONSERVATIVE ORIENTATION OF EDGES

FOR NORMAL VARIABLES: FISHER Z INDEPENDENT TEST

FOR NON-NORMAL CONTINUOUS VARIABLES: CONDITIONAL CORRELATION INDEPENDENCE TEST (RAMSEY, 2014)

FOR DISCRETE VARIABLES: X^2 OR G^2 STATISTICS

VARIABLES OF INTEREST

THE DATASET COMES FROM THE **ALTROCONSUMO** DATABASE AND INCLUDES CHEMICAL AS WELL AS SENSORY VARIABLES AND A GLOBAL SCORE.

FOR EACH YEAR, FROM 2006 TO 2012, ABOUT 280 WINES WERE BOUGHT AND SOME CHEMICAL AND SENSORY CHARACTERISTICS WERE MEASURED.

THE WINES WERE CHOSEN IN ORDER TO REPRESENT THE VARIETY OF ITALIAN VINEYARDS, PRODUCERS AND REGION OF ORIGIN.

FOR EACH YEAR, DIFFERENT VINEYARDS AND PRODUCERS WERE CONSIDERED, SO THAT THE OBSERVATIONS CAN BE CONSIDERED AS INDEPENDENT.

THE SENSORY CHARACTERISTICS WERE EVALUATED WITH THE HELP OF BRESCIA'S **CENTRO STUDI ASSAGGIATORI** USING A PANEL OF EXPERTS BALANCED FOR AGE AND SEX AND THEIR MEDIAN SCORE WAS ATTRIBUTED FOR EACH CHARACTERISTIC TO EACH WINE.

VARIABLES OF INTEREST

THE SIX **CHEMICAL VARIABLES** CONSIDERED HERE ARE:

- WINE'S VERIFIED ALCOHOLIC STRENGTH - (%VOL) (VERIF.ALCOHOL)
- RESIDUAL SUGAR - (G/L) (SUGAR)
- TOTAL ACIDITY - (G/L) (ACIDITY.TOT)
- VOLATILE ACIDITY - (G/L) (ACIDITY.VOL)
- TOTAL SULPHUR DIOXIDE - (MG/L) (SO2.TOT)
- FREE SULPHUR DIOXIDE - (MG/L) (SO2.FREE)

VARIABLES OF INTEREST

THE EIGHTEEN **SENSORY VARIABLES** WERE DIVIDED IN FOUR GROUPS, REPRESENTING :

- VISUAL CHARACTERISTICS (COLOR.INT AND ATTRAENCY)
- OLFACTORY CHARACTERISTICS (OLFACT.INT, FLORAL, FRUITY, SPICY, VEGETAL, OLFACT.CLEAN AND OLFACT.QUAL)
- GUSTATORY CHARACTERISTICS (STRUCTURE, ROUNDNESS, SOURNESS, BITTERNESS, AROM.RICH AND GUSTATORY.HARMONY)
- INTENSE AROMATIC PERSISTENCE (PERSISTENCE, AFTERTASTE.CLEAN AND AFTERTASTE.QUAL)

VARIABLES OF INTEREST

THE OVERALL VARIABLE IS:

- **GLOBAL SCORE OF QUALITY** (GLOBAL.SCORE)
  IT IS THE ALTROCONSUMO'S INDICATOR OF THE OVERALL QUALITY OF THE WINE AND ITS SCORE RANGES FROM 0 (LOWEST QUALITY) TO 100 (HIGHEST QUALITY).

Global.Score	0 - 54	55 - 59	60 - 64	65 - 74	75 - 100
Global.Score.Cat	Very Low	Low	Medium	High	Very High

THREE DESCRIPTIVE VARIABLES WERE TAKEN INTO ACCOUNT

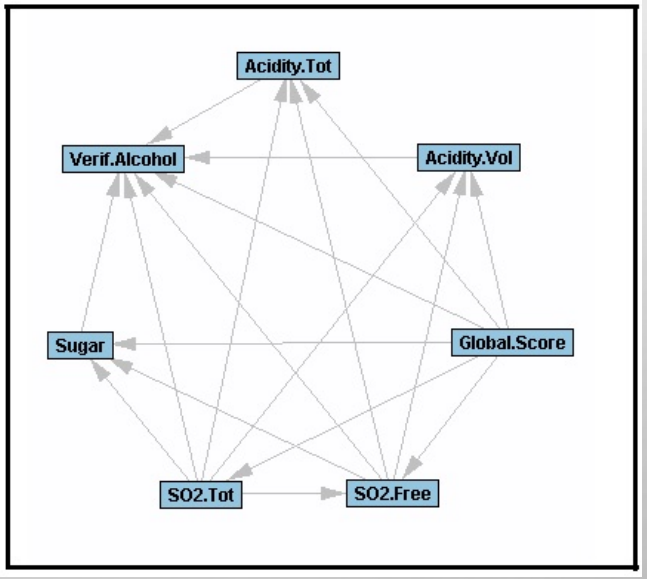
- TYPE OF WINE (TYPE)
- DESIGNATION OF ORIGIN (DENOM)
- REGION OF PRODUCTION (REGION)

Type of wine	Red 1184	White 785	
Designation of origin	DOC 1273	DOCG 269	Other 427
Region of production	North 900	Center 491	South and Islands 578

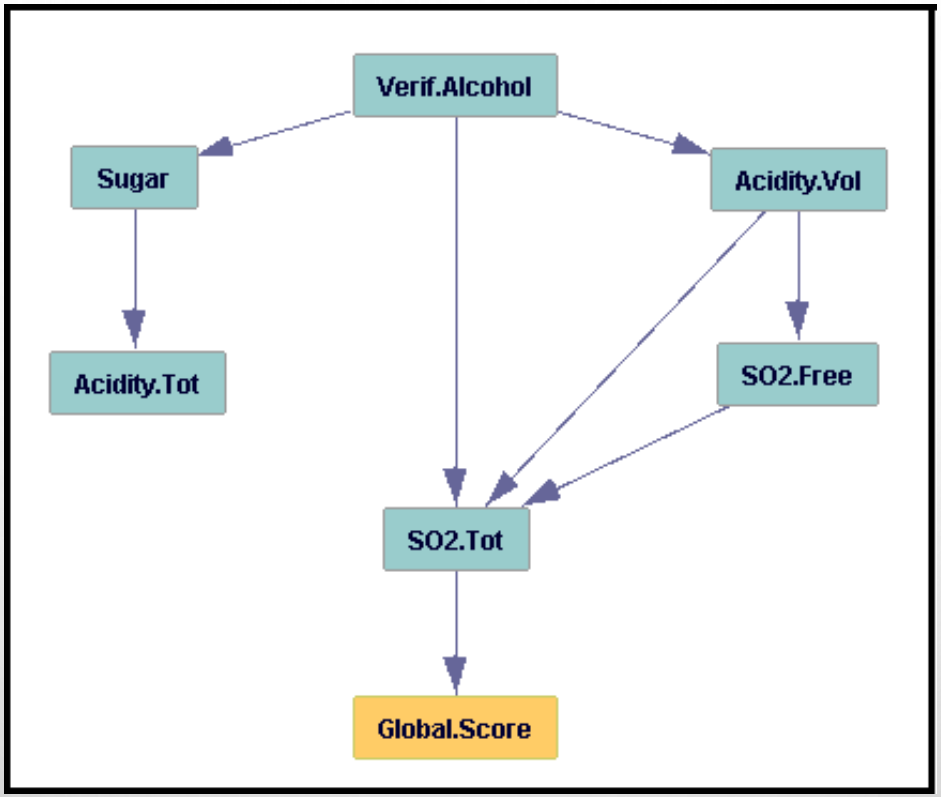
CPDAG FOR CHEMICAL VARIABLES

1969 OBSERVATIONS FROM SIX CONTINUOUS NON NORMAL RANDOM VARIABLES (EXCEPT SO2.TOT) PLUS THE GLOBAL SCORE OF QUALITY WHICH IS CONSIDERED AS CONTINUOUS.

EXPERTS IN THE FIELD SUGGESTED THE CONSTRAINTS VISUALIZED IN THE TWO FIGURES.



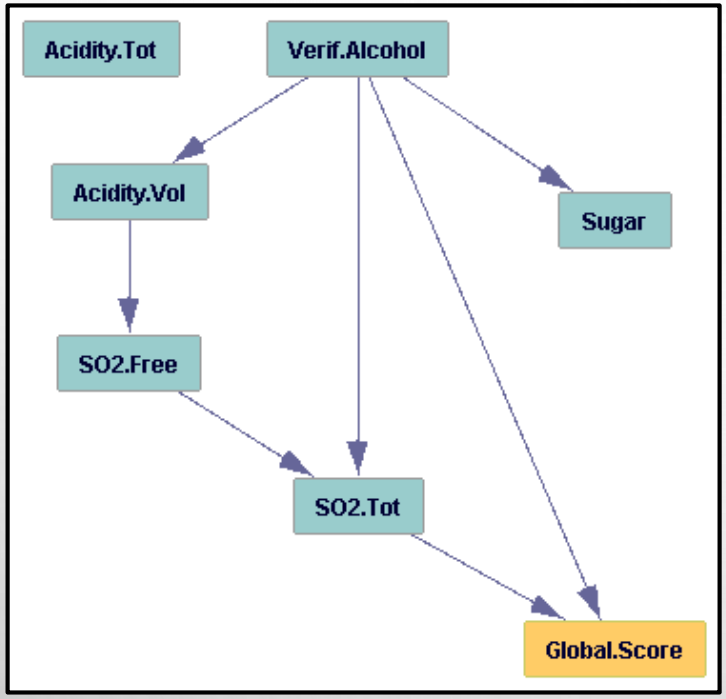
CPDAG FOR CHEMICAL VARIABLES



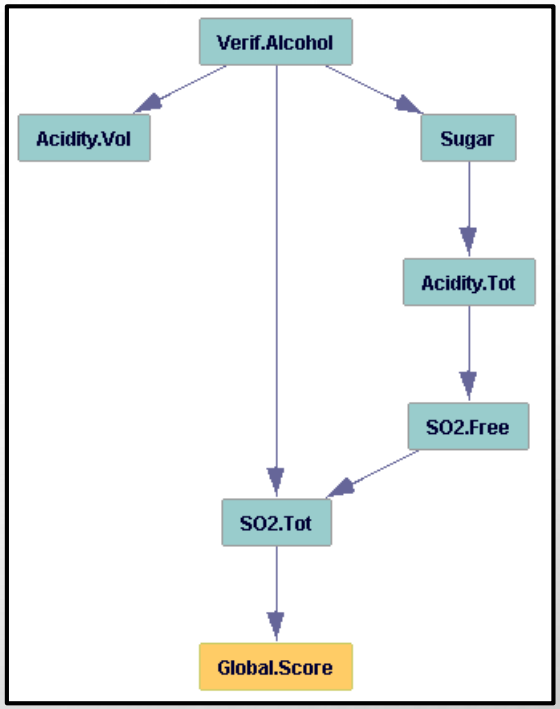
THE RESULTING DAG SHOWS THE RELATIONSHIPS BETWEEN THE CHEMICAL VARIABLES AND HIGHLIGHTS THE ONLY ONE DIRECT CAUSE OF THE **GLOBAL SCORE OF QUALITY**, THAT IS THE **TOTAL SULPHUR DIOXIDE**

CPDAG FOR CHEMICAL VARIABLES

WHITE WINES



RED WINES



FOR BOTH TYPES OF WINES THE **TOTAL SULPHUR DIOXIDE** IS ONE OF THE DIRECT CAUSES OF THE **GLOBAL SCORE OF QUALITY.**

CPDAG FOR SENSORY AND CHEMICAL VARIABLES

1969 OBSERVATIONS FROM 28 CATEGORICAL VARIABLES (2-3-4-5 RESPONSE CATEGORIES).

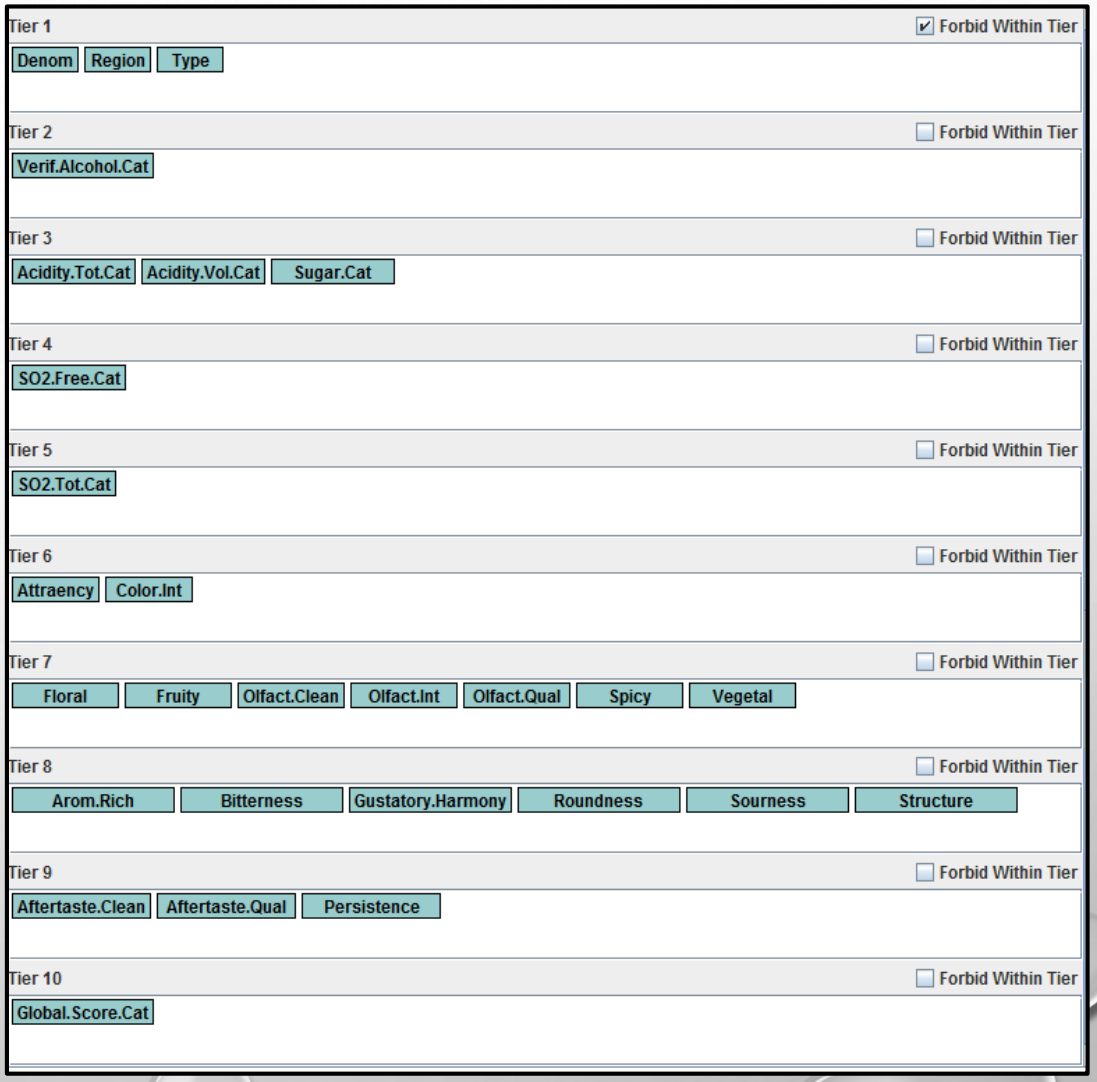
DOMAIN KNOWLEDGE ENTAILS THE CONSTRAINTS WHICH FOLLOW FROM THE TIER ORDERING SHOWN ON THE RIGHT

VISUAL CHARACTERISTICS

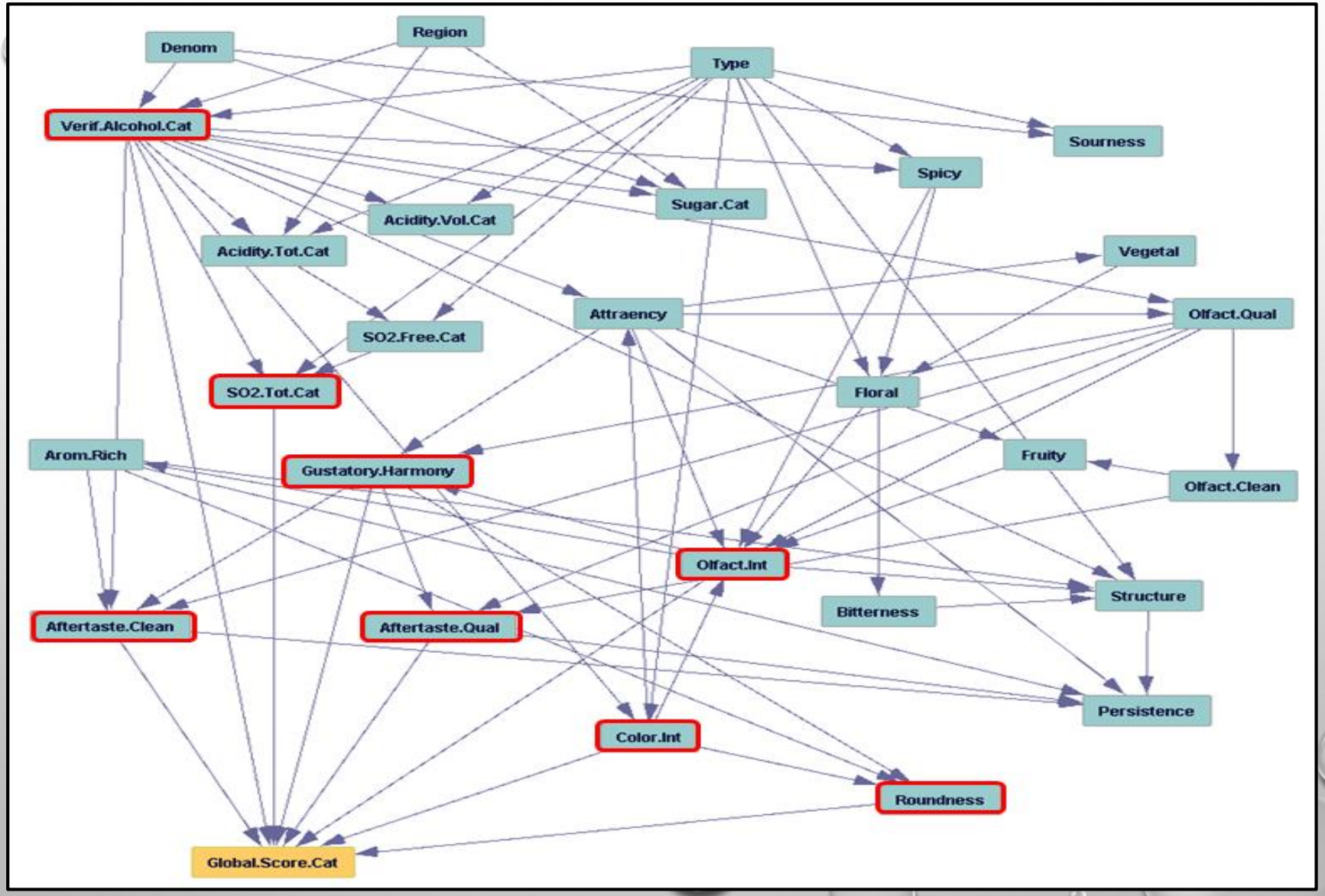
OLFACTORY CHARACTERISTICS

GUSTATORY CHARACTERISTICS

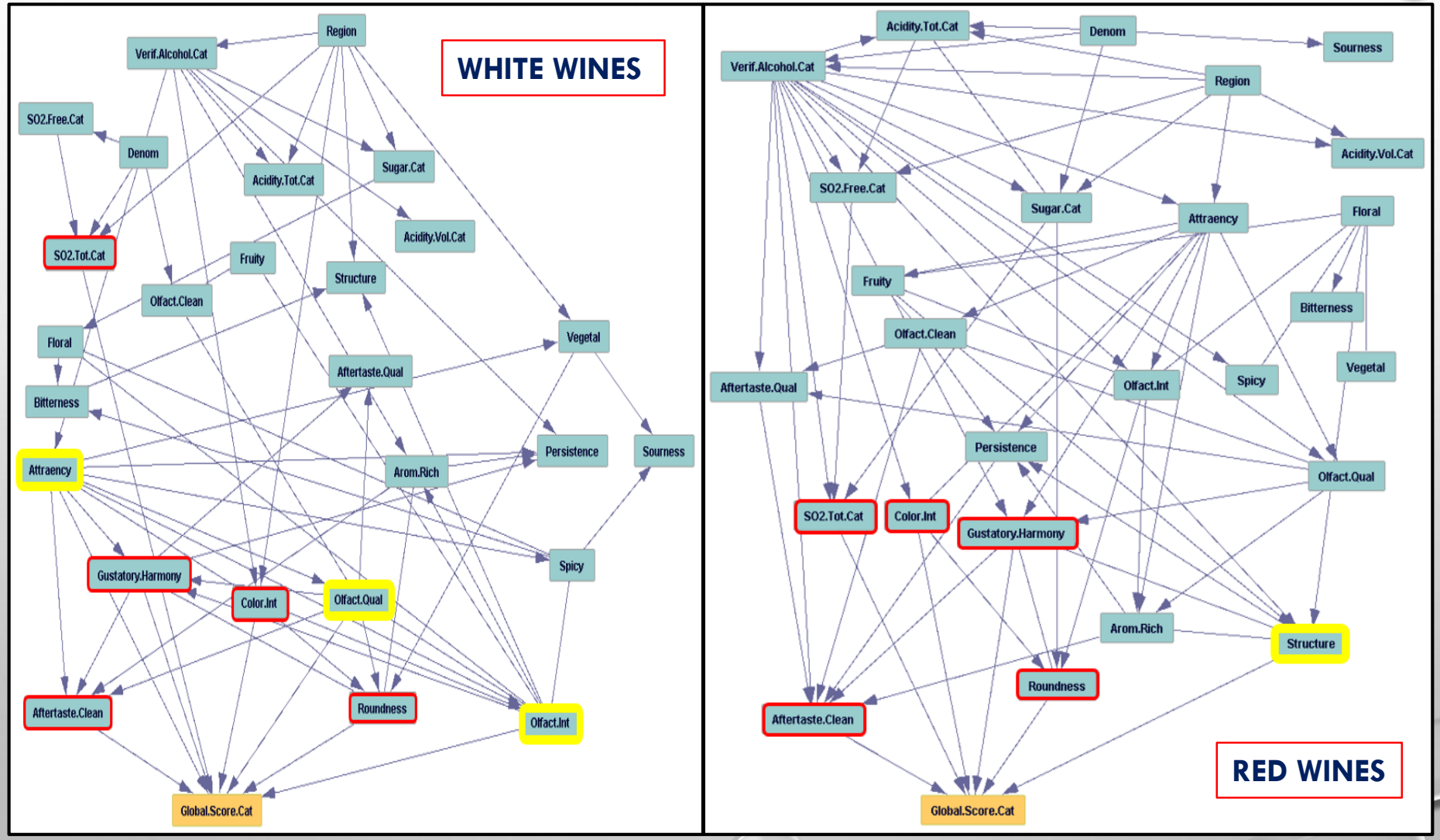
INTENSE AROMATIC PERSISTENCE



CPDAG FOR SENSORY AND CHEMICAL VARIABLES



CPDAG FOR SENSORY AND CHEMICAL VARIABLES



CONCLUSIONS

THE AIM WAS TO IDENTIFY THE FACTORS THAT HAVE A DIRECT INFLUENCE IN DETERMINING THE QUALITY OF THE ITALIAN WINE. TWO KIND OF VARIABLES WERE CONSIDERED, CHEMICAL AND SENSORY VARIABLES. THE ANALYSIS CONSISTED IN THE IDENTIFICATION OF THE EQUIVALENCE CLASS TO WHICH THE TRUE DAG, WHICH VISUALIZES THE CAUSAL RELATIONS BETWEEN THE VARIABLES, BELONGS.

- CONSIDERING THE **CHEMICAL VARIABLES**, THE TOTAL SULPHUR DIOXIDE IS THE ONE PRESENT IN ALL THE CPDAGS WITH A DIRECT EFFECT ON THE GLOBAL SCORE OF QUALITY.
- BETWEEN THE **SENSORY VARIABLES,** THE ONES THAT HAVE A DIRECT EFFECT ON THE GLOBAL SCORE OF QUALITY ARE RELATED TO ALL THE CHARACTERISTICS OF A WINE (APPEARANCE, BOUQUET, FLAVOR AND INTENSE AROMATIC PERSISTENCE).
- CONSIDERING RED AND WHITE WINES SEPARATELY, THE RESULTING CPDAGS CONVEY THE DIFFERENT CHARACTERISTICS OF THE TWO TYPES OF WINE.

REFERENCES

- Guida Vini 2006-2012. Milano: Altroconsumo Edizioni.
- RamseyJ.D., Spirtes P., Zhang J. (2006). Adjacency-Faithfulness and Conservative Causal Inference. *Proceedings of the 22ND Conference on Uncertainty in Artificial Intelligence*, 401-408. Oregon: AUAI Press.
- Ramsey J.D. (2014) A scalable conditional independence test for nonlinear, non-gaussian data. *Arxiv:* http://arxiv.org/abs/1401.5031v2
- Spirtes P., Glymour C., Scheines R. (2000). *Causation, Prediction, and Search*. The MIT Press, 2ND edition.