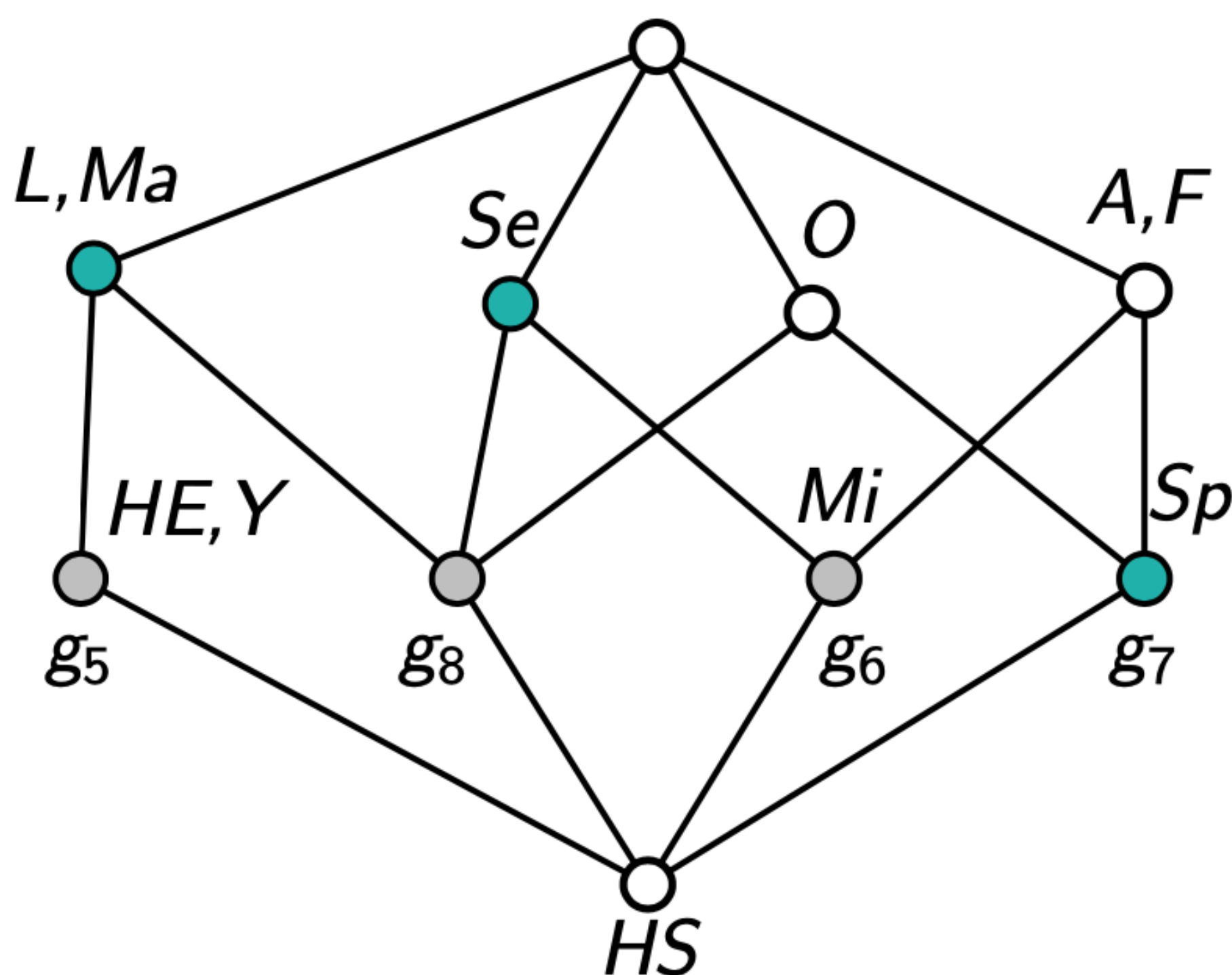


# Power Indices for Attribution of JSM-hypotheses and Formal Concepts

Speaker: Dmitry Ignatov, HSE University



## Abstract

Among the family of rule-based classification models, there are classifiers based on conjunctions of binary attributes. For example, the JSM-method of automatic reasoning (named after John Stuart Mill) was formulated as a classification technique in terms of intents of formal concepts as classification hypotheses. These JSM-hypotheses already represent an interpretable model since the respective conjunctions of attributes can be easily read by decision makers and thus provide plausible reasons for model prediction. However, from the interpretable machine learning (IML) viewpoint, it is advisable to provide decision makers with the importance (or contribution) of individual attributes to the classification of a particular object, which may facilitate explanations by experts in various domains with high-cost errors like medicine or finance. To this end, we use the notion of Shapley value from cooperative game theory, also popular in IML.

In addition to the supervised problem statement, we propose the usage of Shapley and Banzhaf values for ranking attributes of closed sets, namely intents of formal concepts (or closed itemsets). The introduced indices are related to extensional concept stability and are based on counting generators, especially those that contain a selected attribute.

We provide the listeners with theoretical results, basic examples and attribution of JSM-hypotheses and formal concepts by means of Shapley value and some other power indices.



[Registration link](#)