

The background of the slide is a complex, abstract composition. It features a central white rectangular area containing the text. Surrounding this area are various geometric and data-like patterns. On the left, there's a vertical strip with a grid of small '+' symbols and a series of horizontal bars of varying lengths. Above and below the central text, there are large, faint, overlapping shapes that resemble a network or a map. The overall color palette is muted, with shades of brown, beige, and light blue, giving it a technical or academic feel.

4.3. Pattern Evaluation: Null Invariance Measures

In-Lecture Question

Question

DBLP is a computer science bibliography database. Suppose we want to study the correlation between authors Jure Leskovec (JL) and Hector Garcia-Molina (HG) based on their co-authorship, which is shown in the table.

Which of the following measures might be appropriate? Check all that apply.

- ☐ Lift
- ☐ χ^2
- ☐ Cosine
- ☐ Kulczyzyski
- ☐ AllConf

	JL	\neg JL
HG	1	430
\neg HG	133	2.7e6

Question

DBLP is a computer science bibliography database. Suppose we want to study the correlation between authors Jure Leskovec (JL) and Hector Garcia-Molina (HG) based on their co-authorship, which is shown in the table.

Which of the following measures might be appropriate? Check all that apply.

- ☐ Lift
- ☐ χ^2
- ☐ Cosine
- ☐ Kulczyzyski
- ☐ AllConf

	JL	\neg JL
HG	1	430
\neg HG	133	2.7e6

- ☐ **Answers:** Cosine, Kulczyzyski, AllConf
- ☐ **Explanation:** Since both JL and HG have published hundreds of papers and they have only co-authored one paper, JL and HG are negatively correlated. We have $\text{Cosine}(\text{JL}, \text{HG}) = 4.1\text{e-}3$, $\text{Kulc}(\text{JL}, \text{HG}) = 4.9\text{e-}3$, and $\text{Allconf}(\text{JL}, \text{HG}) = 2.3\text{e-}3$, which are small and agree with our observation.
- ☐ However, we have $\text{Lift}(\text{JL}, \text{HG}) = 46.87$ and $\chi^2(\text{JL}, \text{HG}) = 44.79$, which are large due to a large number of null transactions. Therefore, Lift and χ^2 are not appropriate in this case and we should use measures that are null-invariant .