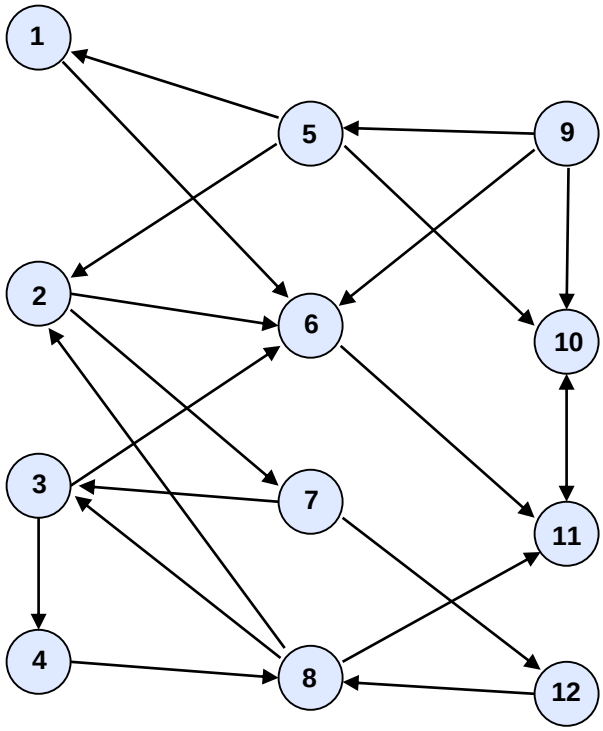
## Task 1: Searching with Lucene (practical)

In this exercise, we use Lucene and its fuzzy retrieval model to search for music files. The web site of the course contains a list of file names, but you can also use your own music library.

- Download Lucene from Apache. Choose the programming language that fits you the best.
- Write a program to read the MP3 file names, create the index, and search for the titles that match your query. You can also use RAMDirectory for a fast implementation (but you need to build the index every time again)
- Extend the basic search with an implementation of the "Did you mean?" function that Google provides. If the query contains spelling mistakes (or is seldom), automatically search with the closest matches of the terms used.
    - Hint: Consider using the SpellChecker of Lucene

## Task 2: Hubs, Authorities, SALSA und PageRank (theoretical)

The following sub-graph of the Internet is given:




In this task, we order the nodes by their hub, authority, and PageRank values

a) We have defined matrices $\mathbf{M}$ and $\mathbf{A}$ for the iterations. In this sub task we use the original HITS algorithm:

$$\boldsymbol{r}^{(t+1)} = \frac{1-\alpha}{N} \cdot \mathbf{1} + \alpha \cdot \mathbf{M} \cdot \boldsymbol{r}^{(t)}$$

$$h^{(t+1)} = \mathbf{A} \cdot a^{(t)}$$

$$a^{(t+1)} = \mathbf{A}^{\mathsf{T}} \cdot h^{(t)}$$

Compute the matrices for the example graph.

b) Write a small program (e.g., with MATLAB, but also works with Excel) that evaluates the fix-point iteration to obtain all results.

c) For the example graph, determine the best hubs, authorities, and the documents with high PageRanks.

d) Apply the SALSA algorithm to the example graph. Does the order change compared to the original HITS algorithm?