Ranking Educational Videos
The Impact of Social Presence

Dimitrios Kravvaris, Katia Lida Kermanidis and Konstantinos Chorianopoulos

Department of Informatics, Ionian University

Objectives

- Create a better ranking procedure for educational videos
- Analyze the impact of social presence

Research

- Educational Videos (from YouTube)
- Content Based Ranking (Cosine Similarity, speaker's transcript)
- The addition of the social weight (likes/(likes+dislikes))
- User Evaluation

Content and Social Approach

- Cosine Similarity $$s = \cos(\vec{q}, \vec{d}) = \frac{\vec{q} \bullet \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\vec{q}}{|\vec{q}|} \bullet \frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$

Content and Social Approach

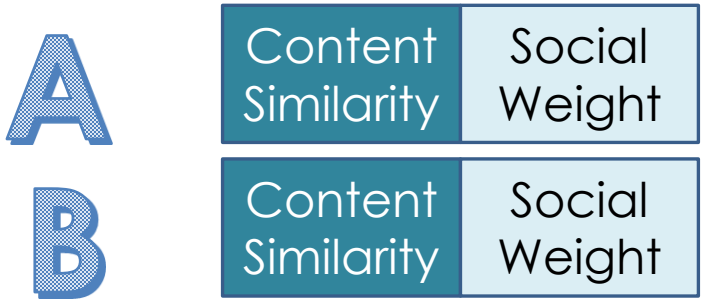
- Cosine Similarity $$s = \cos(\vec{q}, \vec{d}) = \frac{\vec{q} \bullet \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\vec{q}}{|\vec{q}|} \bullet \frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$

- Social Content Similarity $$scs = s(1 + sw) = \cos(\vec{q}, \vec{d})(1 + \frac{likes}{likes + dislikes})$$

Social Content Similarity Analysis

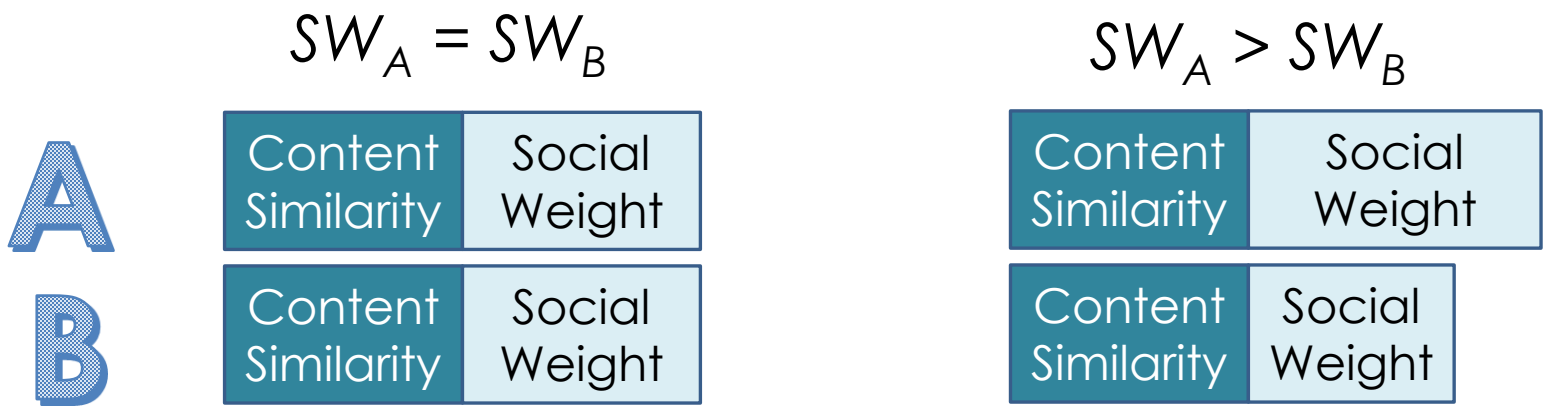
Same Cosine Similarity

$$SW_A = SW_B$$



Social Content Similarity Analysis

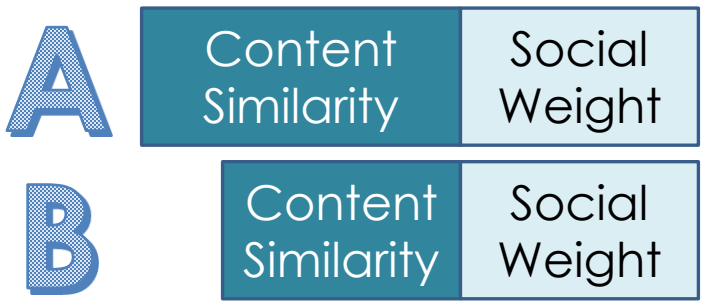
Same Cosine Similarity



Social Content Similarity Analysis

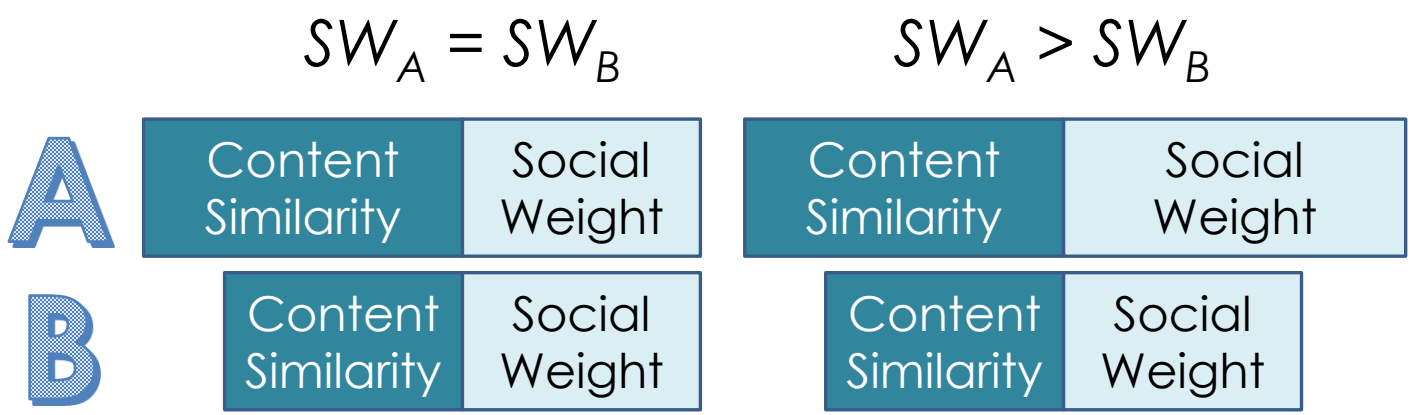
Content Similarity A > Content Similarity B

$$SW_A = SW_B$$



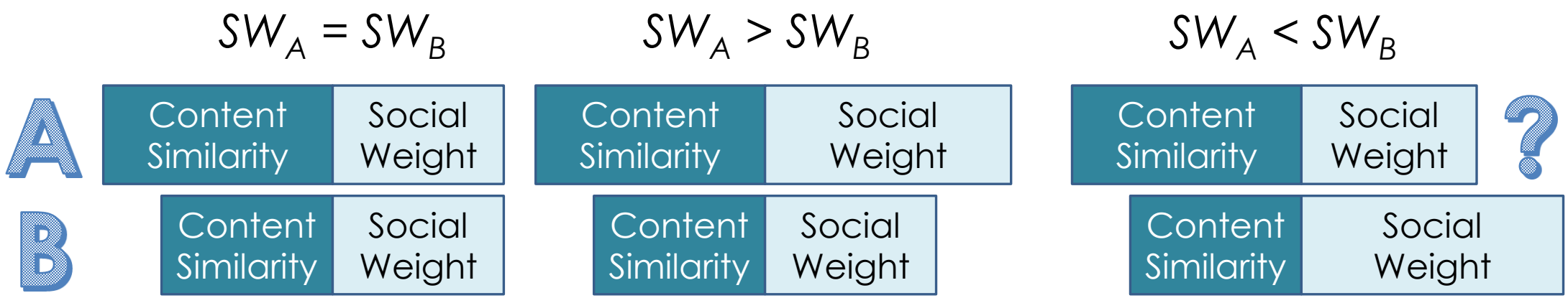
Social Content Similarity Analysis

Content Similarity A > Content Similarity B



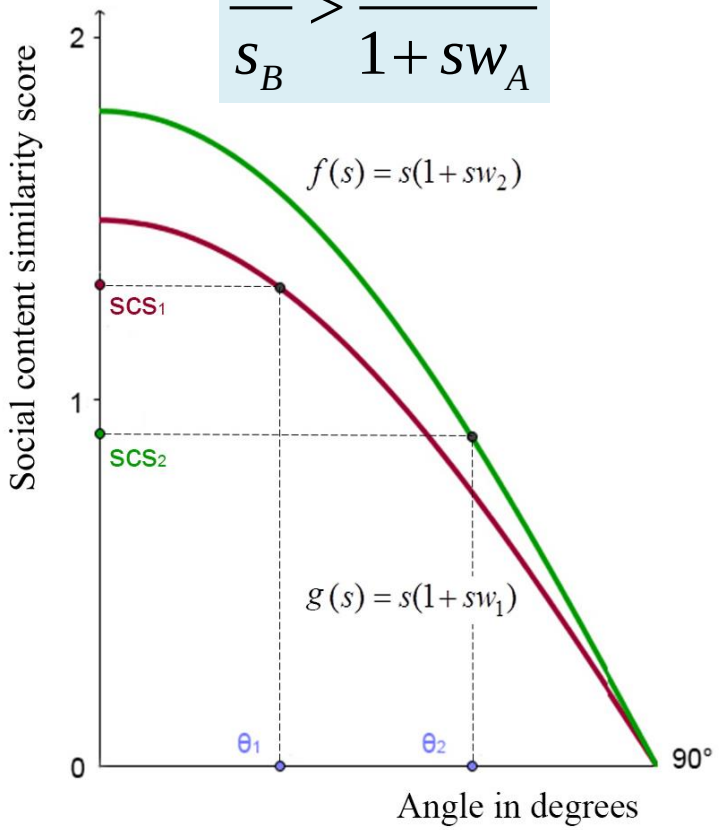
Social Content Similarity Analysis

Content Similarity A > Content Similarity B



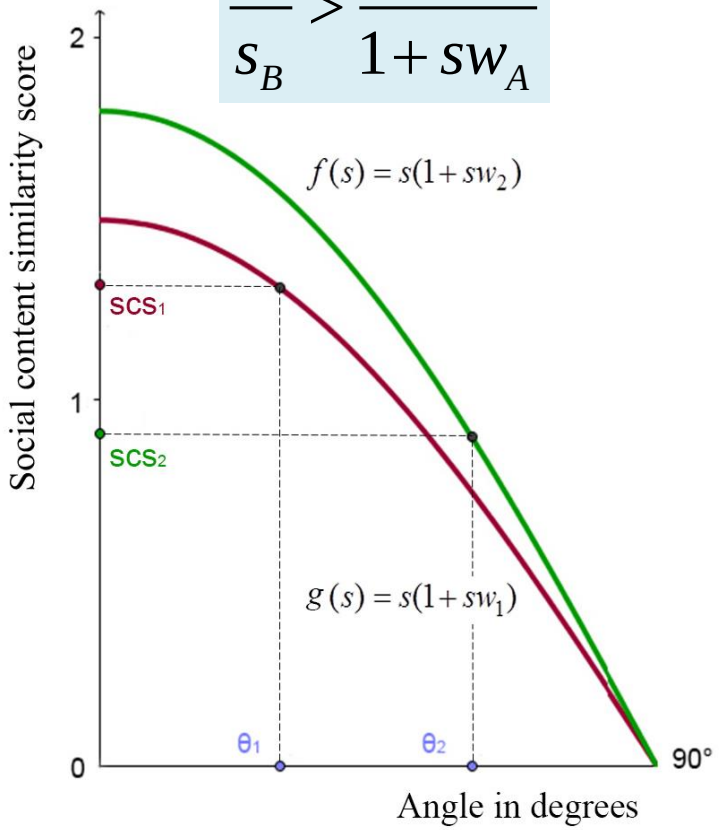
Content Similarity $_A$ > Content Similarity $_B$ AND Social Weight $_A$ < Social Weight $_B$

$$\frac{s_A}{s_B} > \frac{1+sw_B}{1+sw_A}$$

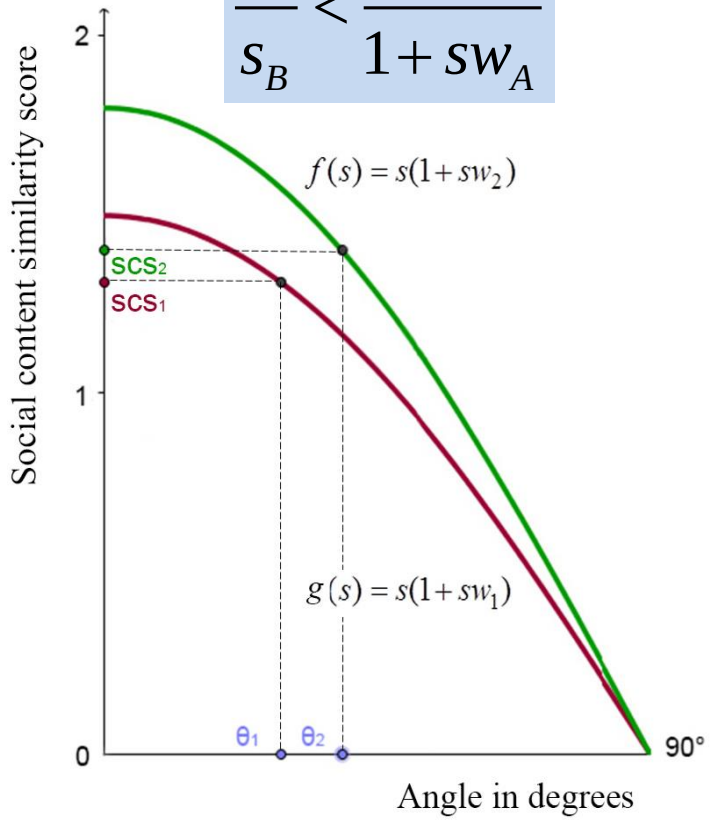


Content Similarity $_A$ > Content Similarity $_B$ AND Social Weight $_A$ < Social Weight $_B$

$$\frac{s_A}{s_B} > \frac{1+sw_B}{1+sw_A}$$

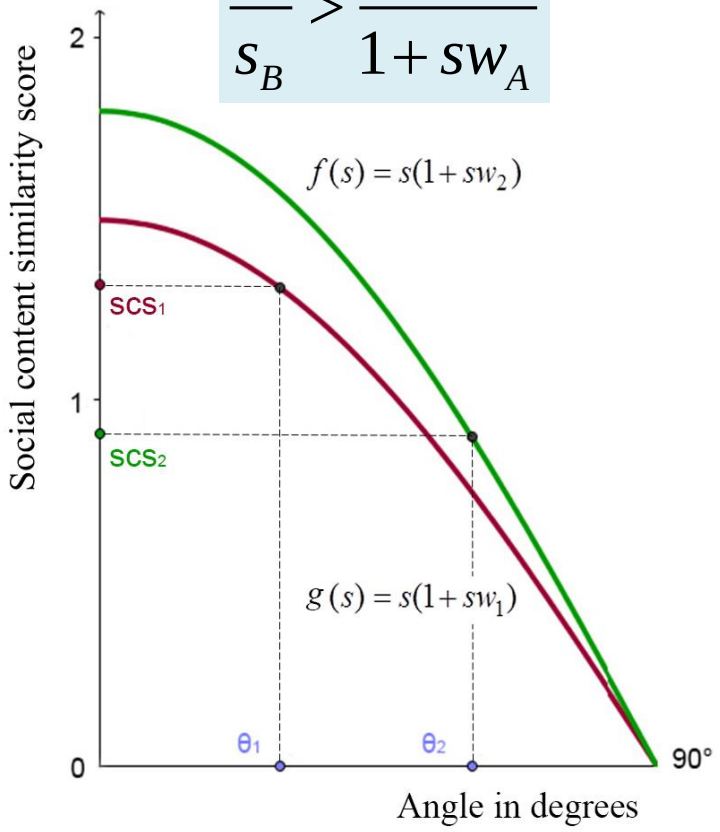


$$\frac{s_A}{s_B} < \frac{1+sw_B}{1+sw_A}$$

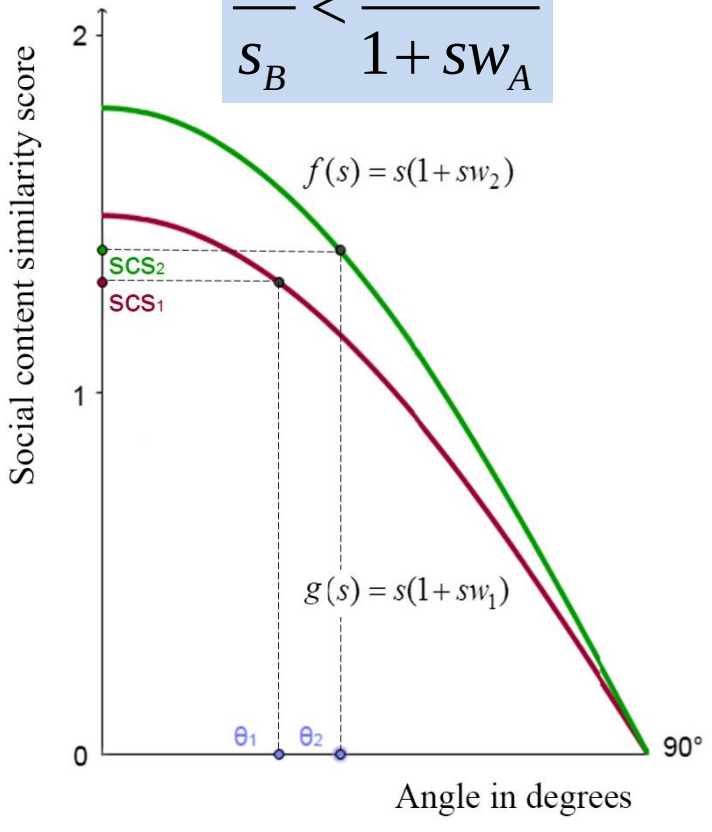


Content Similarity $_A$ > Content Similarity $_B$ AND Social Weight $_A$ < Social Weight $_B$

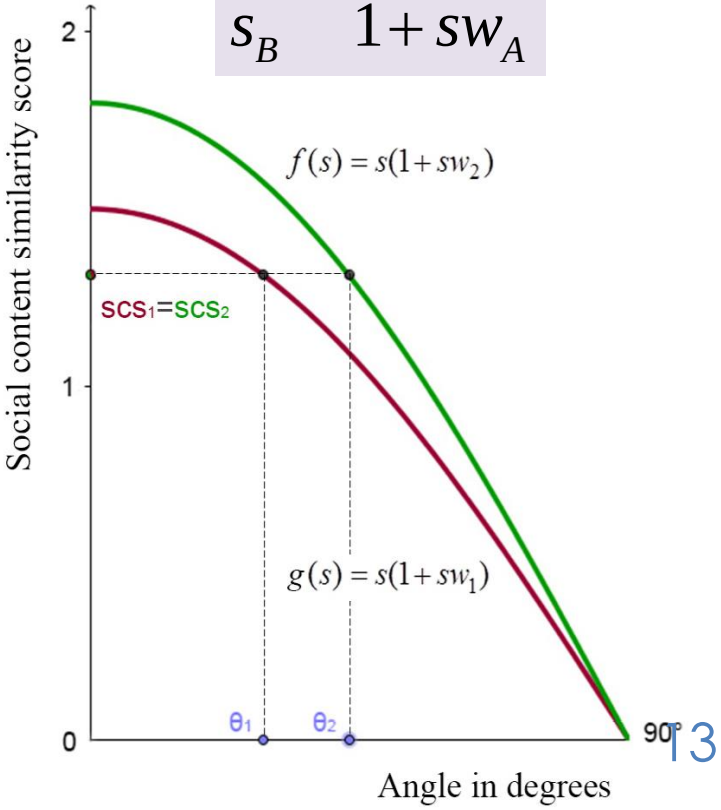
$$\frac{s_A}{s_B} > \frac{1+sw_B}{1+sw_A}$$



$$\frac{s_A}{s_B} < \frac{1+sw_B}{1+sw_A}$$



$$\frac{s_A}{s_B} = \frac{1+sw_B}{1+sw_A}$$



Data

- Video lectures collected from YouTube.
- Searching through the category of Education of YouTube by inserting 40 unique keywords.

  (e.g. computer science, data mining, biology, medicine, statistics, theory, art, social, physics, health, java, analysis, space, network, geography, mathematics)
- 20.830 video lectures were collected among which 1.116 (5.4%) had English transcript.
  - Total duration: 473 hours.
  - Views: over 242 million.

Methodology

- Dataset
  - All the transcripts are processed in Rapidminer according to the following procedures:
    - Tokenize (non letters).
    - Filter Stopwords.
    - Stem English words.
    - Transform cases.

Methodology

- Cluster
- K-means. K=40, which is the number of entries (keywords) used initially as search queries.
  - Simple and flexible algorithm that is easy to understand and explains the clustering outcome.
- Cosine Similarity measure in order to calculate the distance between the objects in our clustering procedure.
  - Suitable technique to be applied to high-dimensional text data, such as the transcripts of video lectures.

Methodology

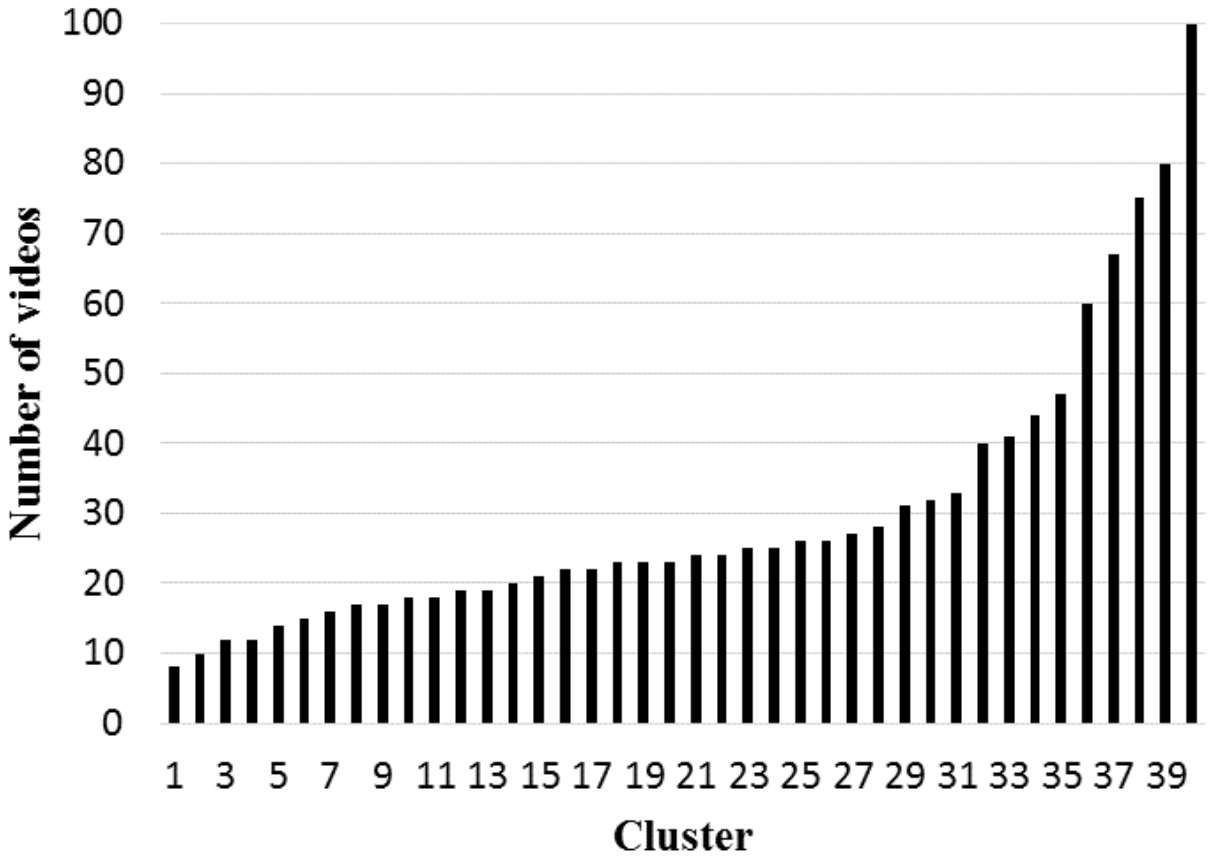
- Data to Similarity
- To every cluster we add a query as a text document and follow the "*Data to Similarity*" procedure in Rapidminer.
- The result is a table with values from all the pair combinations of transcripts including the query based on cosine similarity.
- By ranking them we can see the similarity rank of the transcripts in relation to the search query.

Methodology

- Re-rank
- We add social weight which corresponds to every video lecture and we re-rank them.
- The new ranking has this time a social character and we can study the differences between the previous and the new one.
- We examined the theoretical cases that appear according to the first part of our study.

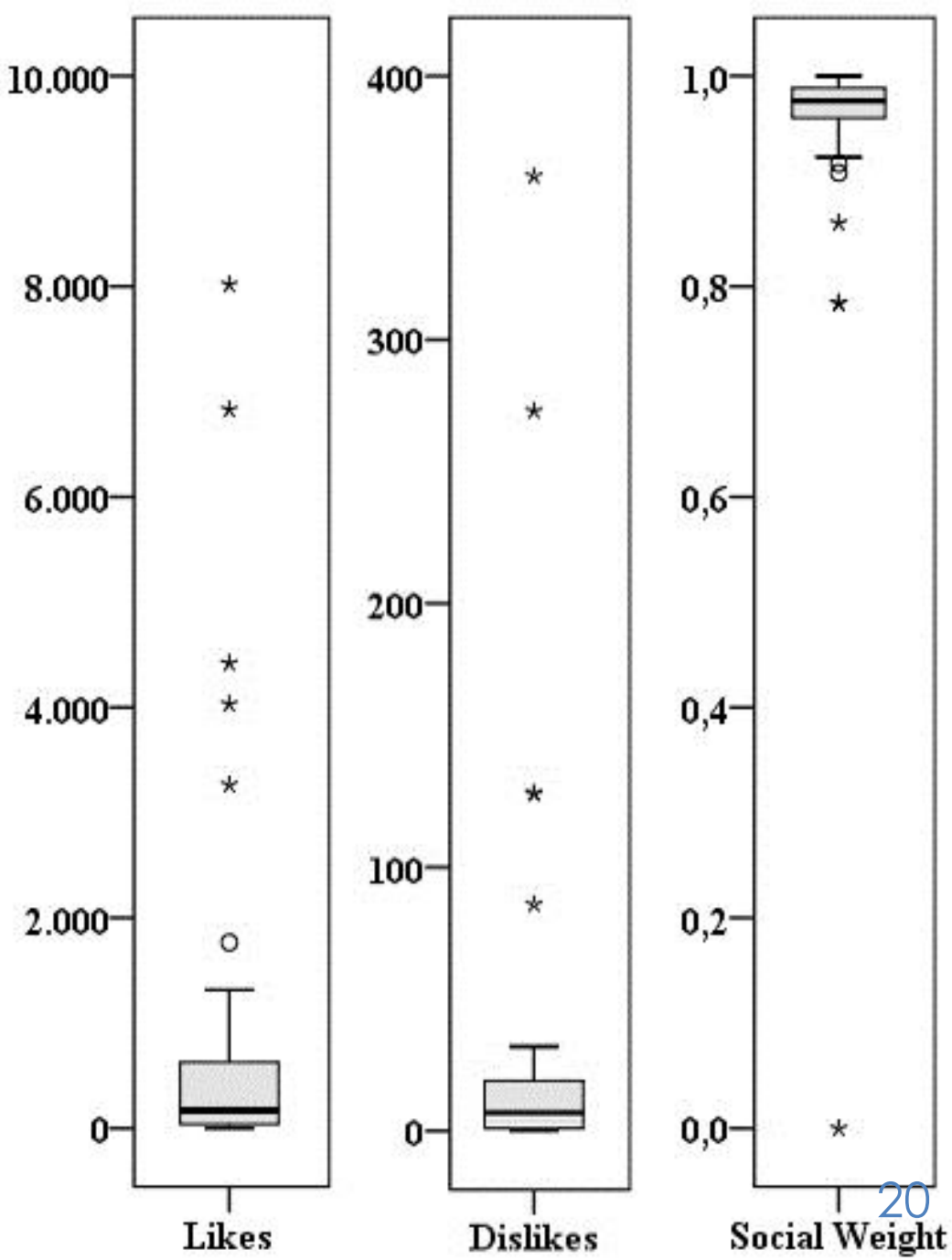
Experimental Results

- Clusters
- 40 clusters were created, which on average contain 29 video lectures each.



Experimental Results

- Social Weight
- The social weight values is on average equal to 0.96
- There are some extreme cases where there is no *like* but there are *dislikes*
- Cases where there are only likes on a video lecture are more frequent (13%).

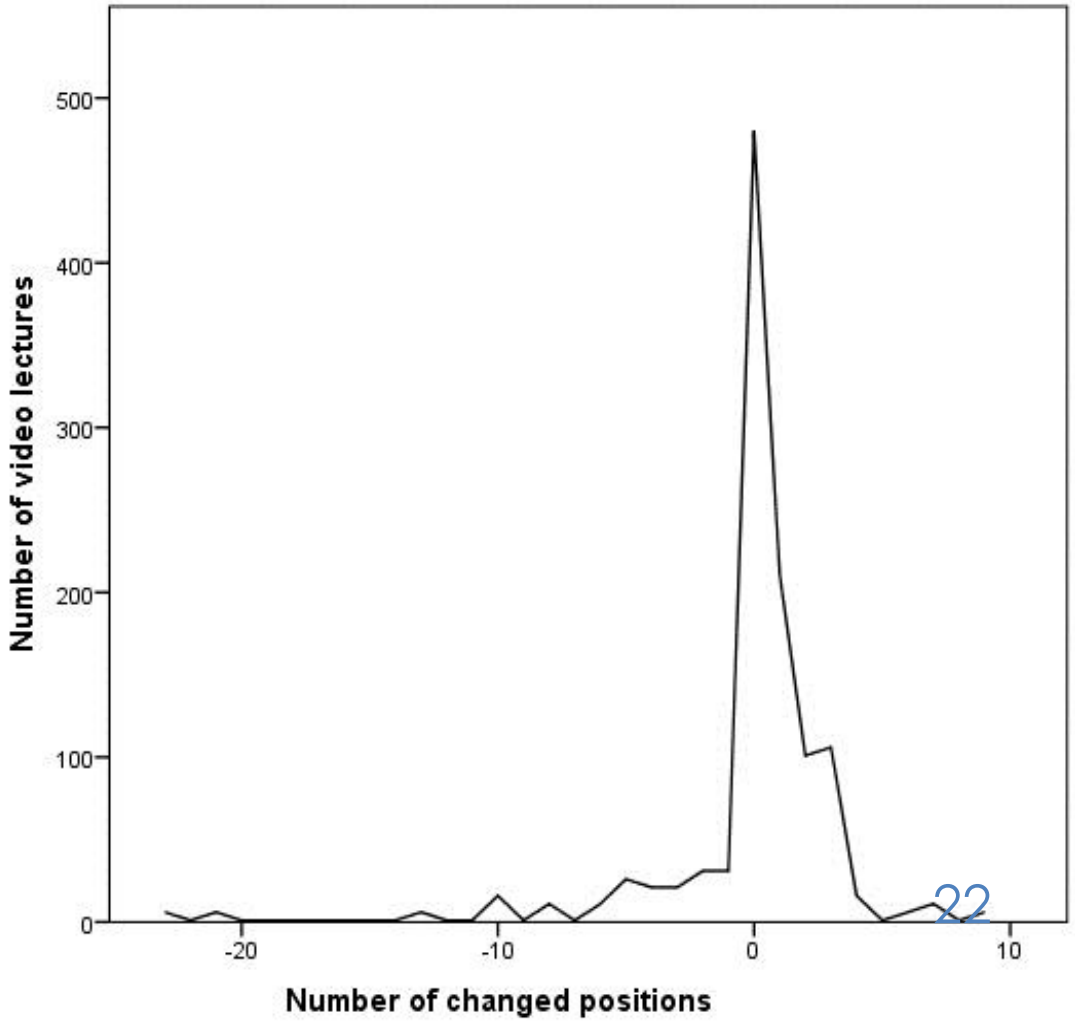


Experimental Results

- We find out that the ranking of a great 43% of the video lectures has not changed.
- But 57% of the video lectures under examination have changed, which shows the dynamics that social weight has on the ranking of the results.
- Social weight had a positive effect 41% of the video lectures while only 16% of them went down to a lower position in the ranking.

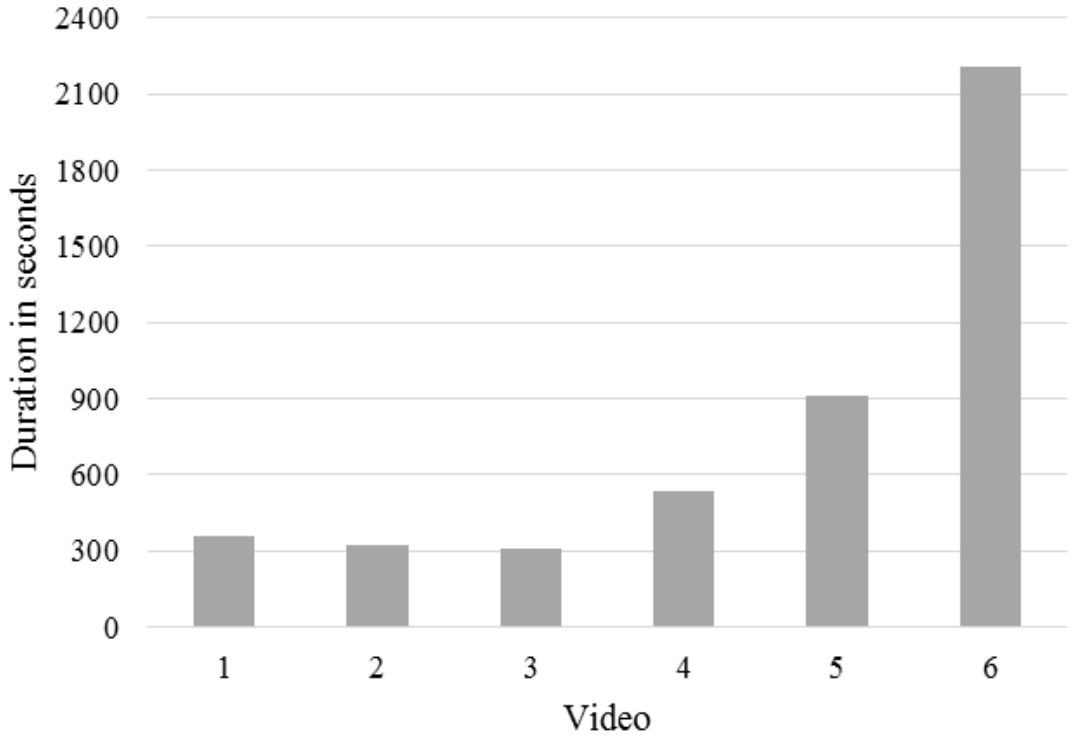
Experimental Results

- Changed Positions
- The changes in the ranking of the video lectures are usually 1 to 4 positions up or down, which covers about 81.74% of the video lectures that have changed order.
- There is the case, however, where a video lecture has gone down 23 positions. In this case we find out that the *sw*=0 and more specifically there is 1 *dislike* and no *likes*.



User Evaluation

- Data
- The data were collected from 15 users-rankers who belong to different age groups and have different interests.
- We used online questionnaires which contained 6 videos from the keyword "*database*" clusters.
- The total duration of the videos was 77 minutes



User Evaluation

- Methodology
- The videos were presented to each ranker in a random order.
- Each user had to rank the videos in the order they would like them to be presented after a possible search of the word “database” on the YouTube.
- Furthermore, they were asked if they watched the whole videos or stopped watching them at some point
- They were also asked to describe both the positive and the negative features each video had, in their opinion, which helped to rank them.
- Finally, there was an interview with five of the rankers which aimed at recording their attitude towards the videos they liked the most or not at all.

User Evaluation

- Methodology
- Our research question is to find out which of the two aforementioned ranking methods is closer to the users' ranking.
- We assume that every user's ranking is correct.
- We compare the ranking of every ranker to the content similarity video ranking and the social-content similarity video ranking respectively. We chose to use the Mean Average Precision (MAP) measure for this quantitative comparison.

User Evaluation

- Experiment

	Ranking Order					
	1st	2nd	3rd	4th	5th	6th
Content Similarity	1	2	3	4	5	6
Social-Content Similarity	1	2	3	5	4	6

	Ranking Order					
	1st	2nd	3rd	4th	5th	6th
user1	2	1	3	5	4	6
user2	2	1	3	5	4	6
user3	2	1	3	5	4	6
user4	2	1	4	3	5	6
user5	2	1	3	5	6	4
user6	2	1	5	3	4	6
user7	2	1	5	6	4	3
user8	2	1	5	3	4	6
user9	2	1	5	3	6	4
user10	2	1	5	3	6	4
user11	2	1	5	3	6	4
user12	2	1	5	4	3	6
user13	2	1	3	5	4	6
user14	2	1	3	5	4	6
user15	2	1	5	3	6	4

User Evaluation

- Results
  - We notice that the social-content similarity method is most of the times close to the users' ranking at satisfying percentages.
  - The social-content similarity method predicts the users preference by 27.39% while the content similarity by 19.23%.

	Mean Average Precision	
	Content Similarity	Social-Content Similarity
user1	33.33%	52.50%
user2	33.33%	52.50%
user3	33.33%	52.50%
user4	26.67%	16.67%
user5	33.33%	41.67%
user6	16.67%	26.67%
user7	0.00%	20.00%
user8	16.67%	26.67%
user9	0.00%	0.00%
user10	0.00%	0.00%
user11	0.00%	0.00%
user12	29.17%	16.67%
user13	33.33%	52.50%
user14	33.33%	52.50%
user15	0.00%	0.00%

Conclusion

- The majority of the changes in position does not affect the new ranking to such a degree that it would bring a less relevant to the content video lecture to the top positions
- A low social weight can cause a video lecture to go down to a much lower position, which indicates the effect of the users' opinions, even on video lectures that are at the top positions based on their content relevance.

Conclusion

- The user evaluation experiment confirmed that the social-content similarity method is more reliable than content similarity method, with a percentage that reaches 52.5% that the videos will be ranked in the same order as they will be ranked by the users.
- It is also confirmed that the content of video is considered to be more important than other characteristics.
- It was found out that the users express their positive opinion (*like*) for the videos more easily than their negative opinion (*dislike*).

Ranking Educational Videos
The Impact of Social Presence

Thank You