Spoken Lecture Summarization by Random Walk over a Graph Constructed with Automatically Extracted Key Terms

Author: Yun-Nung Chen, Yu Huang , Ching-Feng Yeh, and Lin-Shan Lee

Speaker: Hung-Yi Lee

National Taiwan University

Outline

- Introduction
- Graph-based summarization approach
  - A spoken document is transformed into a graph structure
    - Nodes: sentences in a spoken document
    - Edge weight: topical similarities of sentences
  - Random walk is used to select indicative sentences
  - all sentences in a document can be jointly considered
- Experiments
- Conclusion

Introduction – Extractive Summarization (1/2)

- Extractive speech summarization
  - Select the indicative sentences in a spoken document
  - Cascade the sentences to form a summary
  - The number of sentences selected as summary is decided by a predefined ratio

Introduction – Extractive Summarization (2/2)

- Each sentence S in a spoken document d is given an *importance score* I(S,d)
  - Select the indicative sentences based on I(S,d)

sentence term

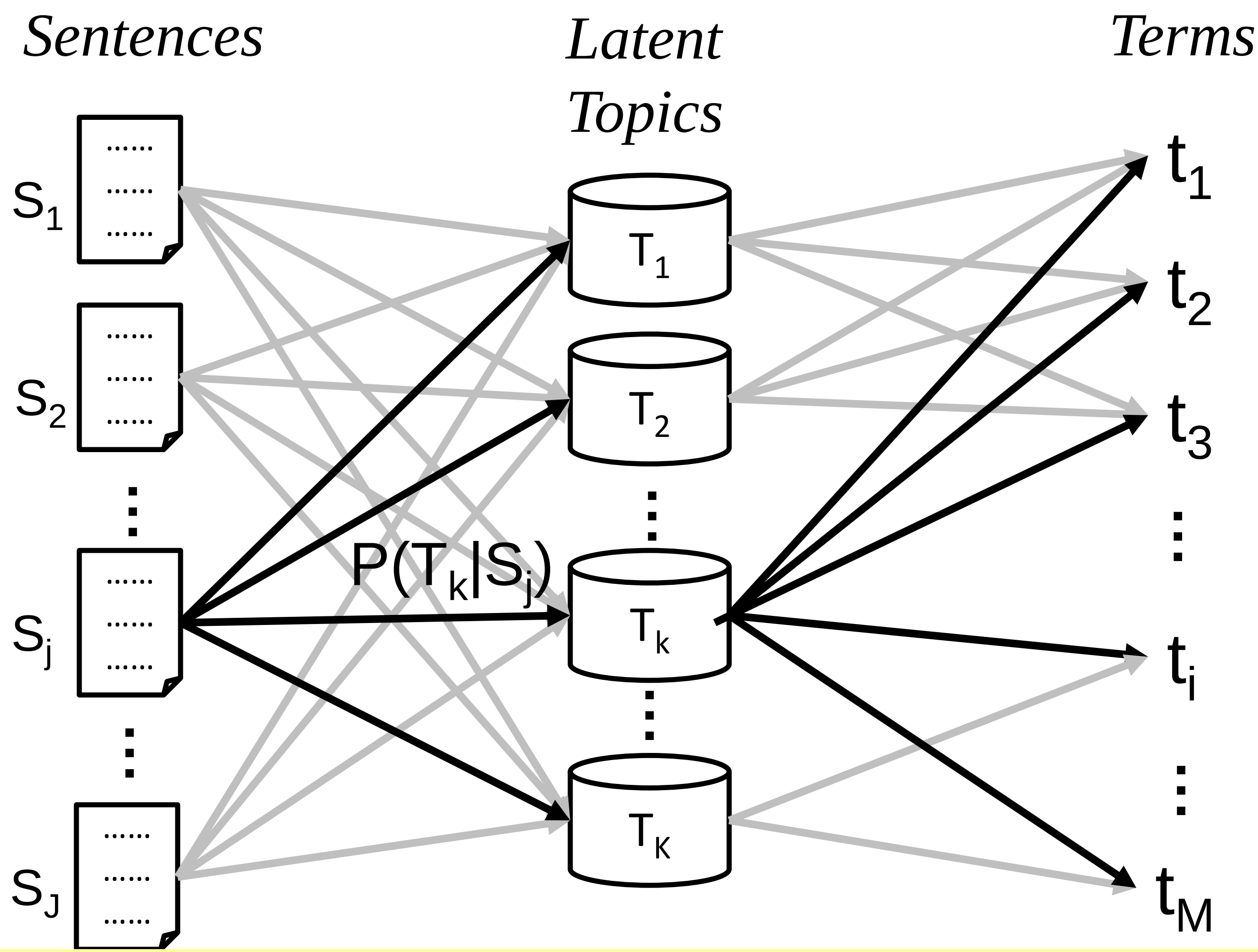
$$S = t_1 t_2 \ldots\ldots t_i \ldots\ldots t_n$$

$$\mathrm{I}(\mathrm{S}, d) = \sum_{i=1}^{n} \left[\boxed{s(t_i, d)} \cdots\cdots \right] + \cdots\cdots$$

Importance score

term statistical measure

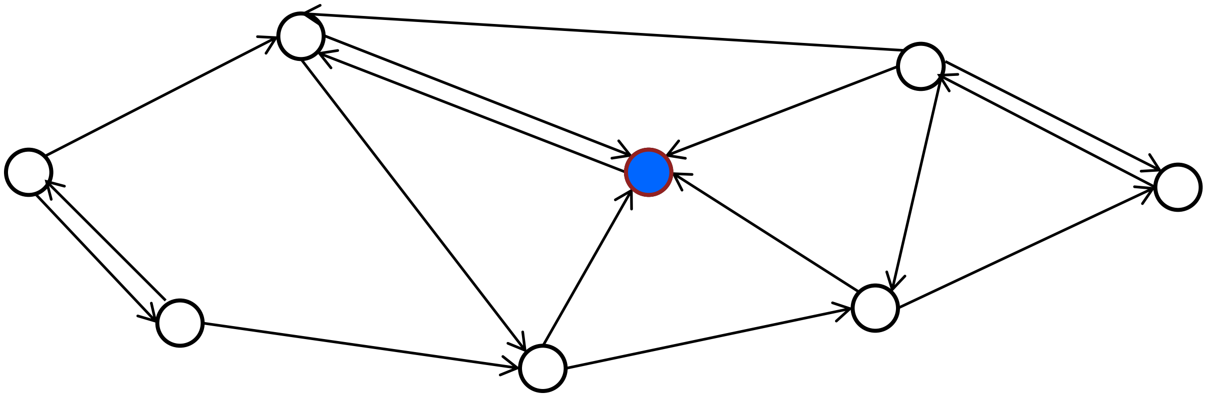
Introduction – PLSA



$P(T_k|S_j)$: weight of latent topic T_k for sentence S_j

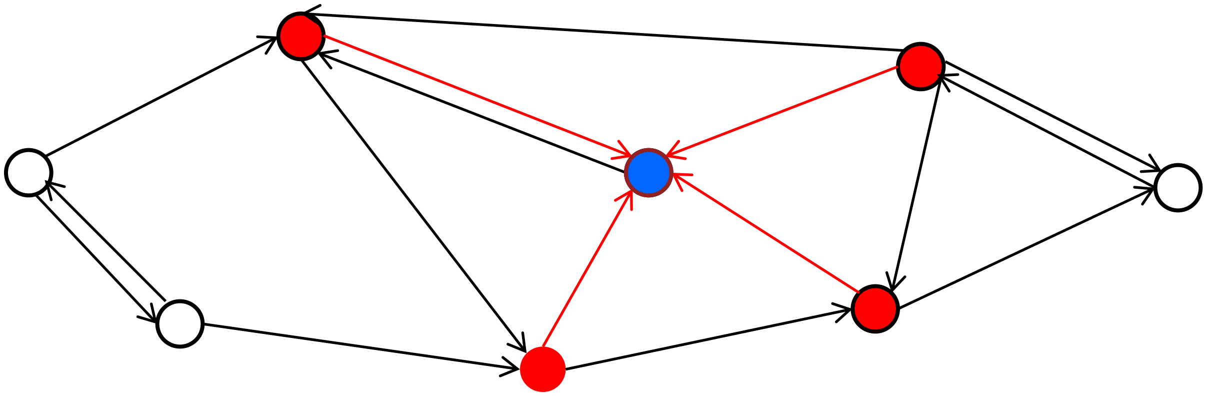
Proposed Approach (1/2)

- Basic idea
  - Not only the sentences with high importance score based on statistical measure should be considered as indicative sentence



Proposed Approach (1/2)

- Basic idea
  - Not only the sentences with high importance score based on statistical measure should be considered as indicative sentence
  - But the sentences topically similar to the indicative sentences should also be considered as indicative



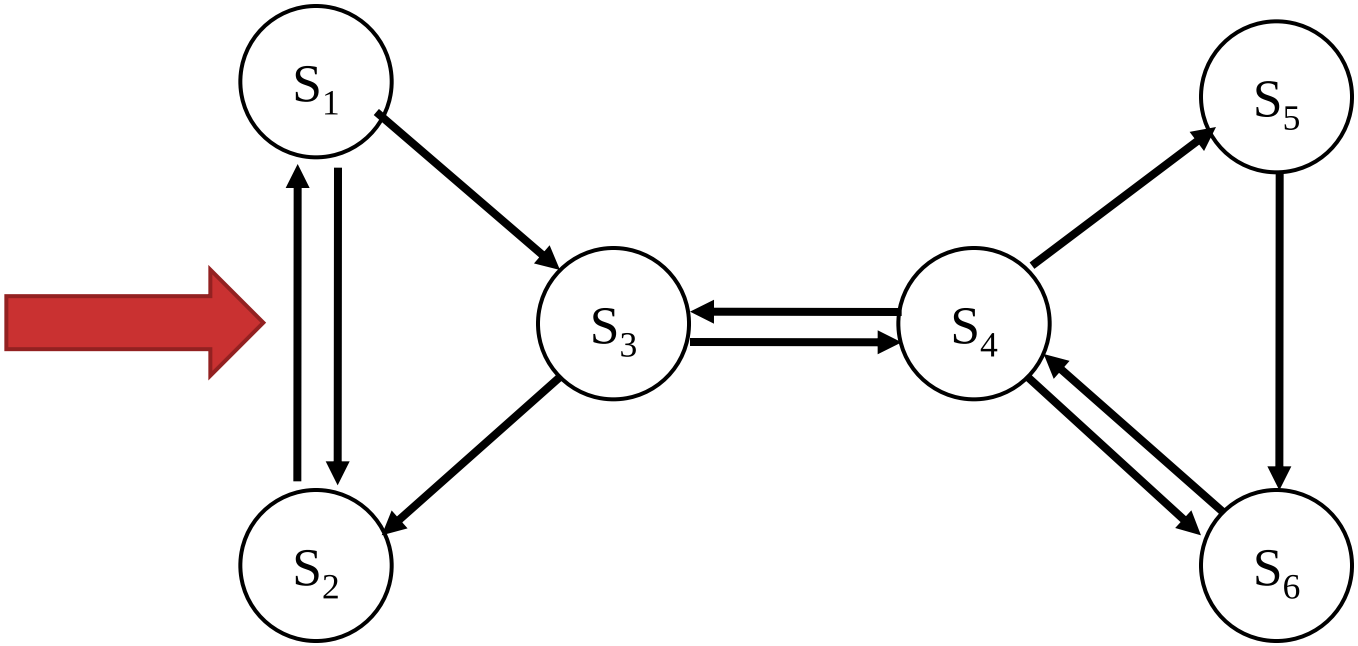
Proposed Approach (2/2)

- Graph-based approach
  - Sentences in a spoken document are nodes on a graph, and topical similarities of sentences are weights of edges.
  - Use random walk to obtain new scores for summary selection
  - → all sentences in the document can be jointly considered rather than individually.

Graph Construction (1/2)

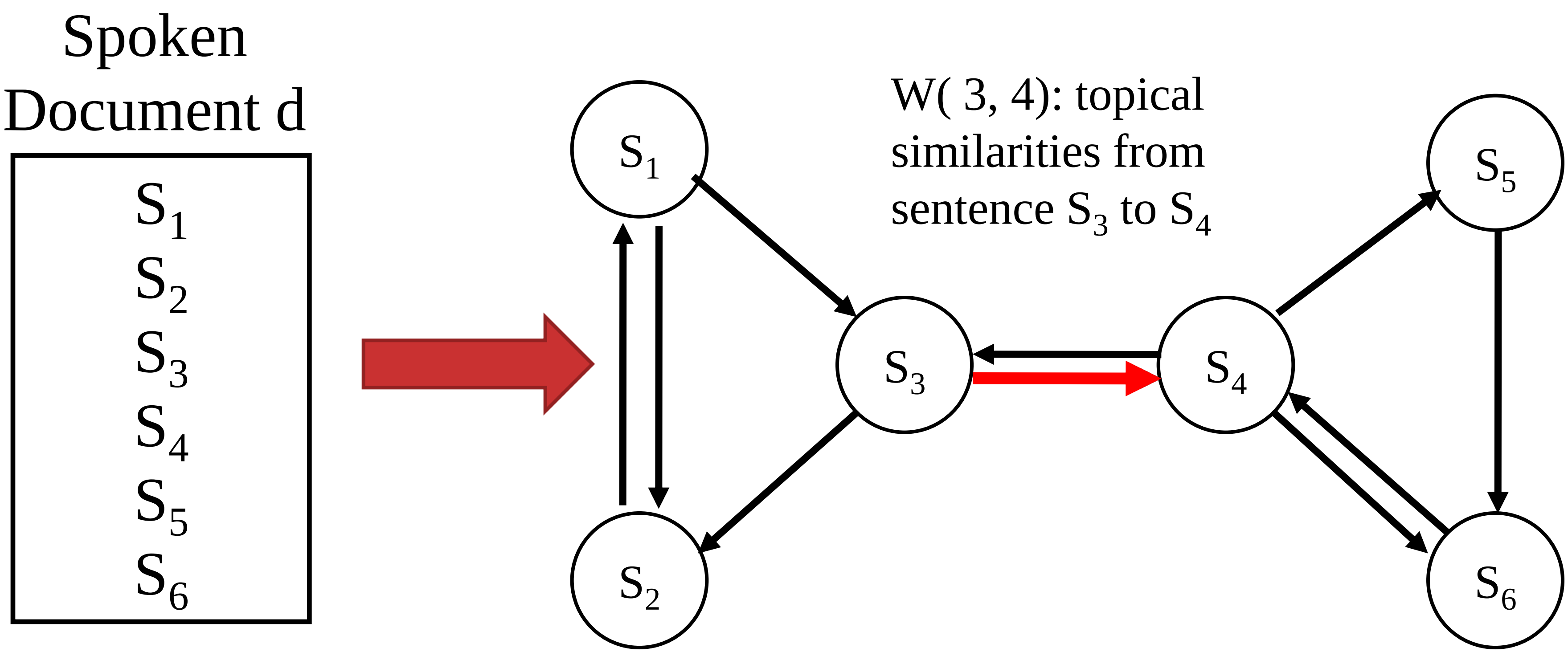
Spoken
Document d

S_1
S_2
S_3
S_4
S_5
S_6



Each sentence S_i in the spoken document d is a node on the graph.

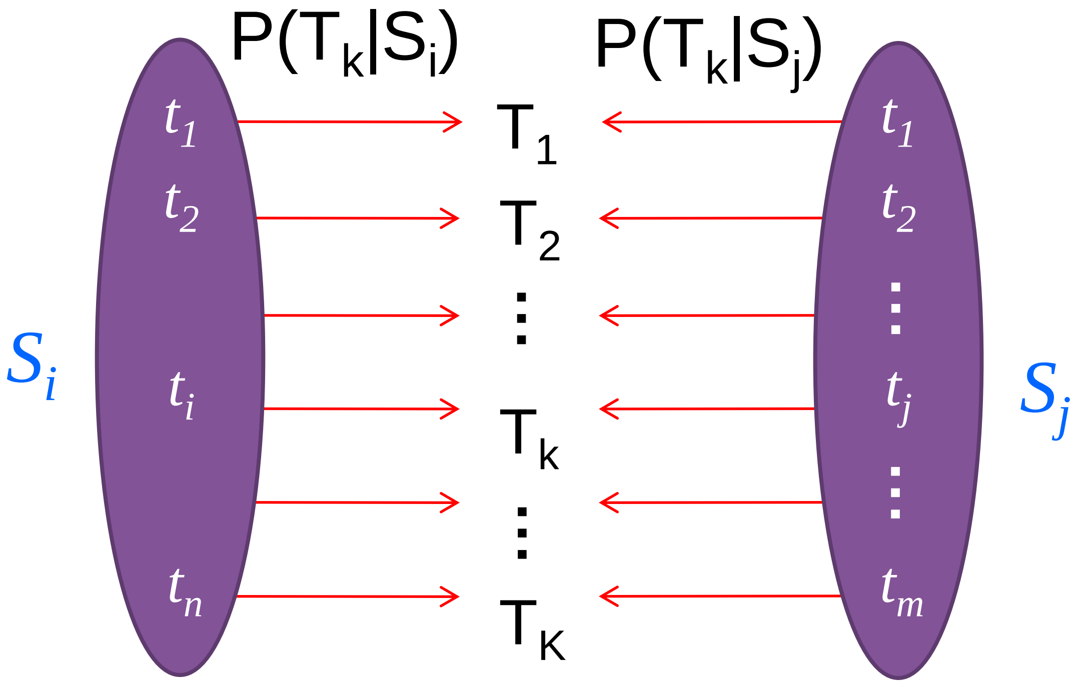
Graph Construction (1/2)



W(i , j) ($S_i \rightarrow S_j$):
Topical similarity from sentence S_i to S_j
(based on *PLSA latent topics* of sentences)

Graph Construction (2/2)
- Topical Similarities

- Topical Similarity from sentences S_i to S_j
  - Edge weight W(i , j) (sentence $S_i \rightarrow$ sentence S_j)

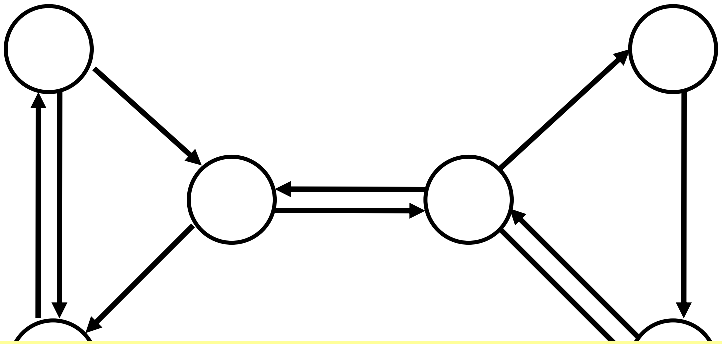


◆ W(i , j): evaluated by the latent topic similarities of sentences S_i to S_j based on PLSA model

Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$\mathrm{G}(i) = (1-\alpha)\mathrm{I}(\mathrm{S}_i, d) + \alpha \sum_{S_j \in in(i)} \mathrm{G}(j)\hat{\mathrm{W}}(j,i)$$



◆ G(i) for sentence S_i would be a new importance score for summary selection

Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = \boxed{(1-\alpha)}\boxed{I(S_i, d)} + \boxed{\alpha}\boxed{\sum_{S_j \in in(i)} G(j)\hat{W}(j,i)}$$

The original importance score of node S_i (weighted by 1-α)

Scores propagate from other nodes to node S_i (weighted by α)

Mathematical Formulation

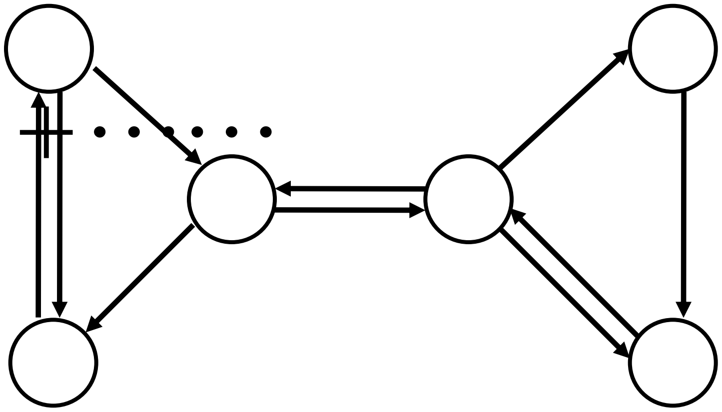
Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha)\boxed{I(S_i, d)} + \alpha \sum_{S_j \in in(i)} G(j)\hat{W}(j,i)$$

$$I(S,d) = \sum_{i=1}^{n} [s(t_i, d)\cdots\cdots] + \cdots\cdots$$

Importance score

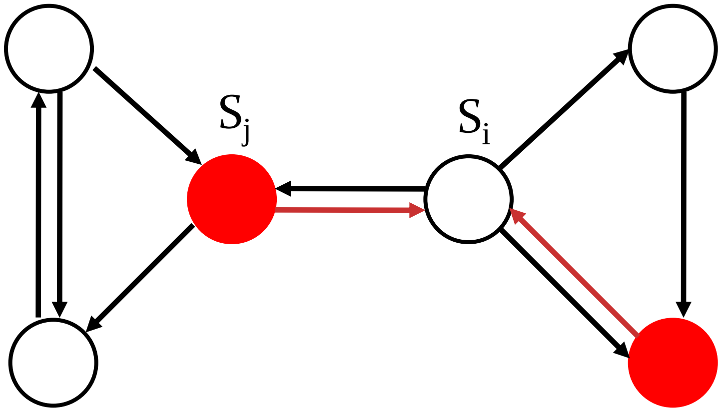
term statistical measure



Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

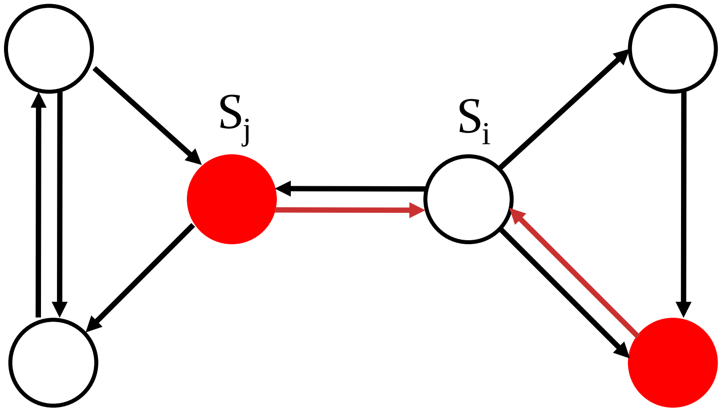
$$\mathrm{G}(i) = (1-\alpha)\mathrm{I}(\mathrm{S}_i, d) + \alpha \sum_{\boxed{S_j \in in(i)}} \mathrm{G}(j)\hat{\mathrm{W}}(j,i)$$



Mathematical Formulation

Find a set of new scores based on graph structure
{G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha)\,I(S_i, d) + \alpha \sum_{S_j \in in(i)} \boxed{G(j)}\boxed{\hat{W}(j,i)}$$

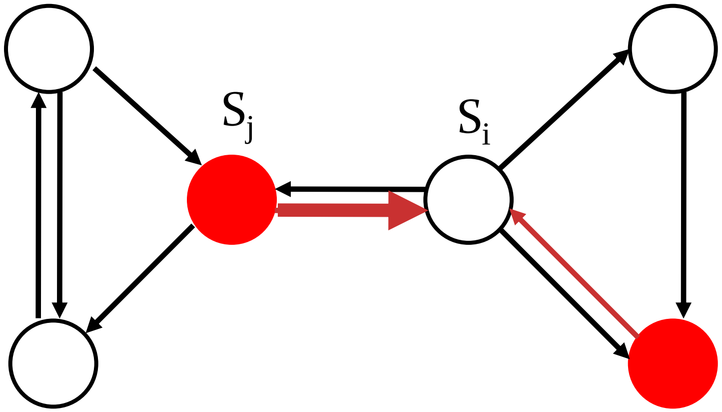


Mathematical Formulation

Find a set of new scores based on graph structure
{G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \boxed{\hat{W}(j,i)}$$

$$\hat{W}(j,i) = \frac{\boxed{W(j,i)}}{\sum_{S_k \in out(j)} W(j,k)}$$



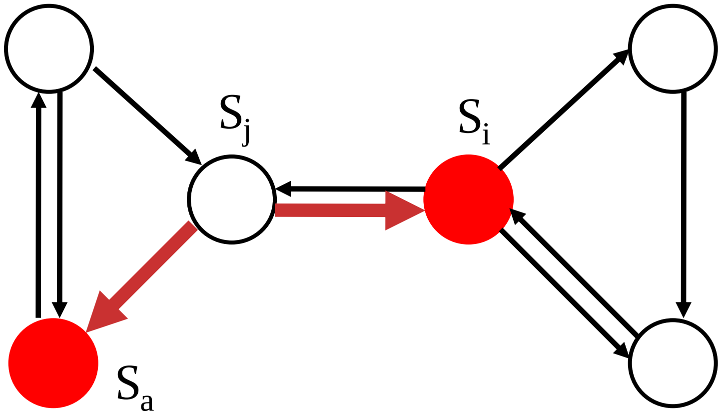
Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \boxed{\hat{W}(j,i)}$$

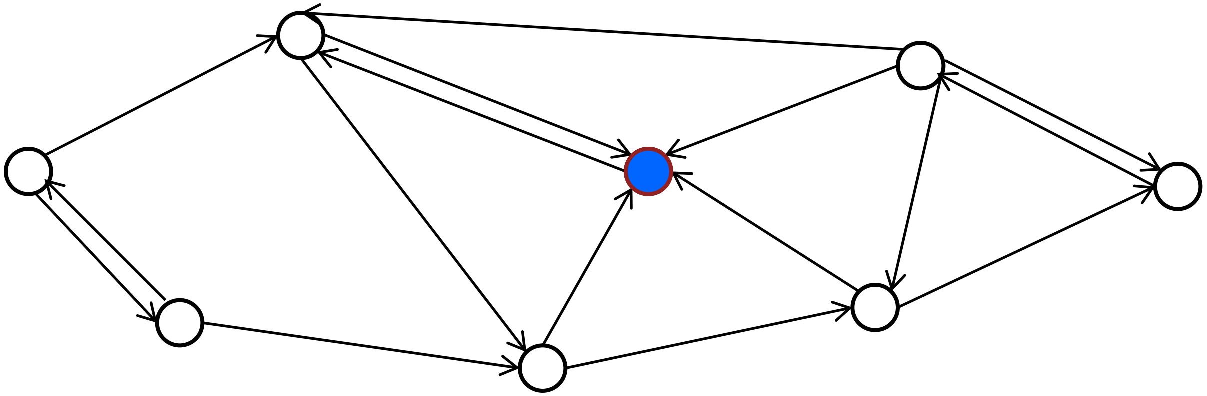
$$\hat{W}(j,i) = \frac{W(j,i)}{\boxed{\sum_{S_k \in out(j)} W(j,k)}}$$

The scores propagate from a node to all other nodes <u>sums to unity</u>.



Mathematical Formulation

$$\mathrm{G}(i) = (1-\alpha)\mathrm{I}(\mathrm{S}_i, d) + \alpha \sum_{S_j \in in(i)} \mathrm{G}(j)\hat{\mathrm{W}}(j,i)$$



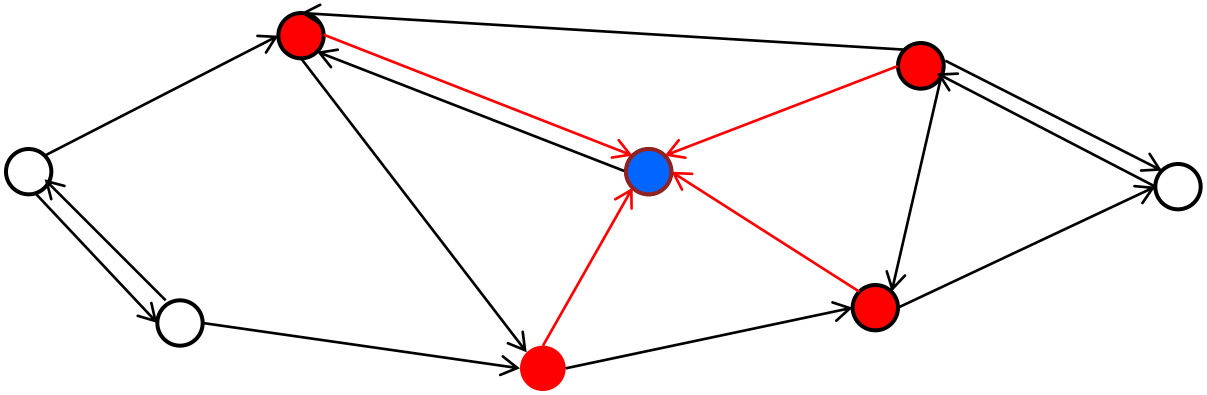
◆ G(i) can obtain higher score when

1) $\mathrm{I}(\mathrm{S}_i,\mathrm{d})$ is high.

2) More sentences topically similar to S_i

Mathematical Formulation

$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \hat{W}(j, i)$$



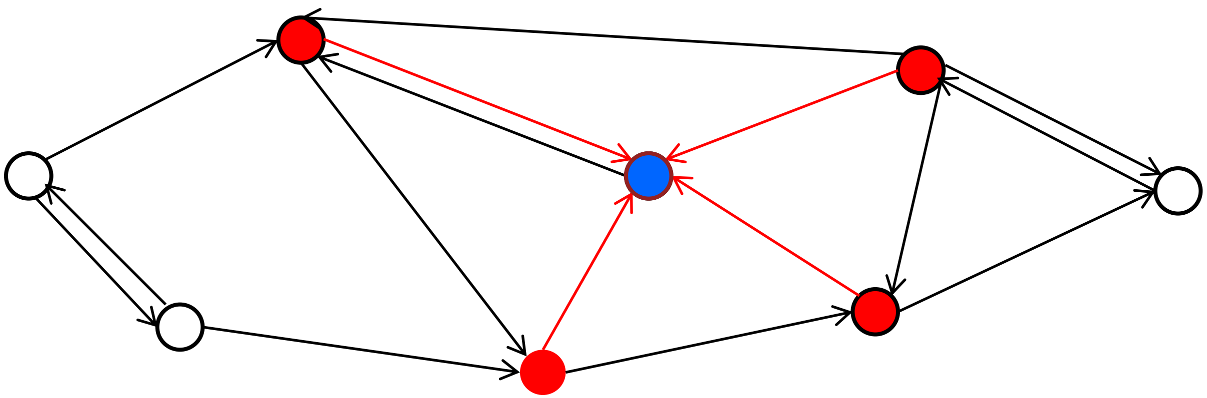
◆ G(i) can obtain higher score when

1) $I(S_i,d)$ is high.

2) More sentences topically similar to S_i

Mathematical Formulation

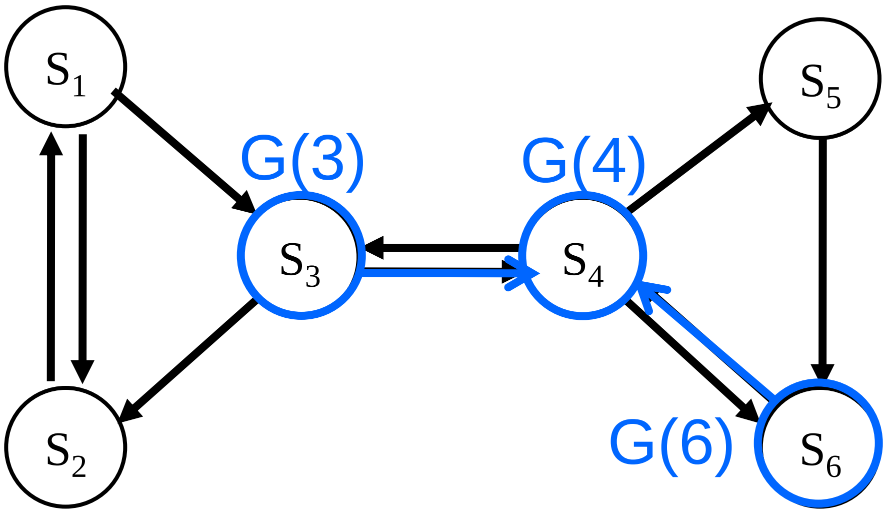
$$G(i)=(1-\alpha)I(S_i,d)+\alpha\sum_{S_j\in in(i)}G(j)\hat{W}(j,i)$$



- All sentences in the documents are considered jointly
- Rather than individually

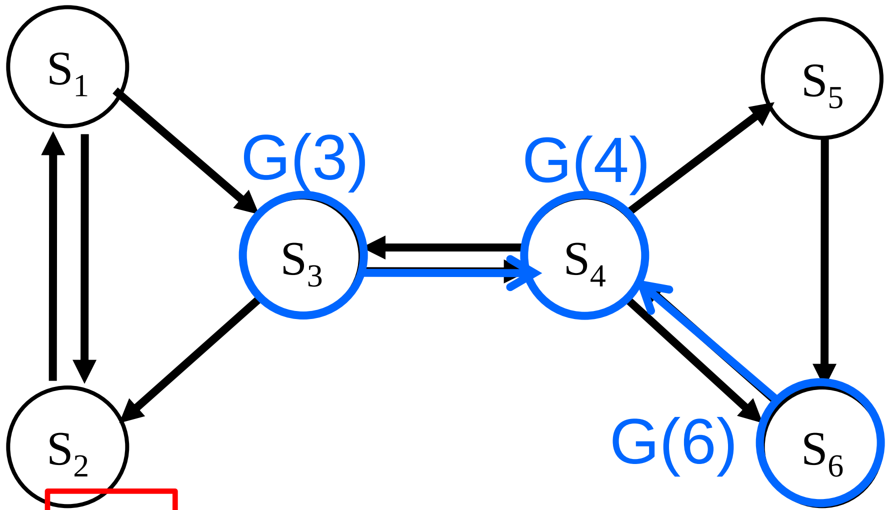
Mathematical Formulation – an Example

Find G(1), G(2), G(3), G(4), G(5), G(6) such that



$$\boxed{G(4)} = (1-\alpha)\boxed{I(S_4, d)} + \alpha \boxed{G(6)}\boxed{\hat{W}(6,4)} + \alpha \boxed{G(3)}\boxed{\hat{W}(3,4)}$$

Mathematical Formulation – Equations to be solved



Find G(1), G(2), G(3), G(4), G(5), G(6) such that

$$G(1) = (1-\alpha)\mathrm{I}(S_1, d) + \alpha G(2)\hat{W}(2,1)$$

$$G(2) = (1-\alpha)\mathrm{I}(S_2, d) + \alpha G(3)\hat{W}(3,2)$$

$$G(3) = (1-\alpha)\mathrm{I}(S_3, d) + \alpha G(1)\hat{W}(1,3) + \alpha \boxed{G(4)}\hat{W}(4,3)$$

$$G(4) = (1-\alpha)\mathrm{I}(S_4, d) + \alpha G(6)\hat{W}(6,4) + \alpha \boxed{G(3)}\hat{W}(3,4)$$

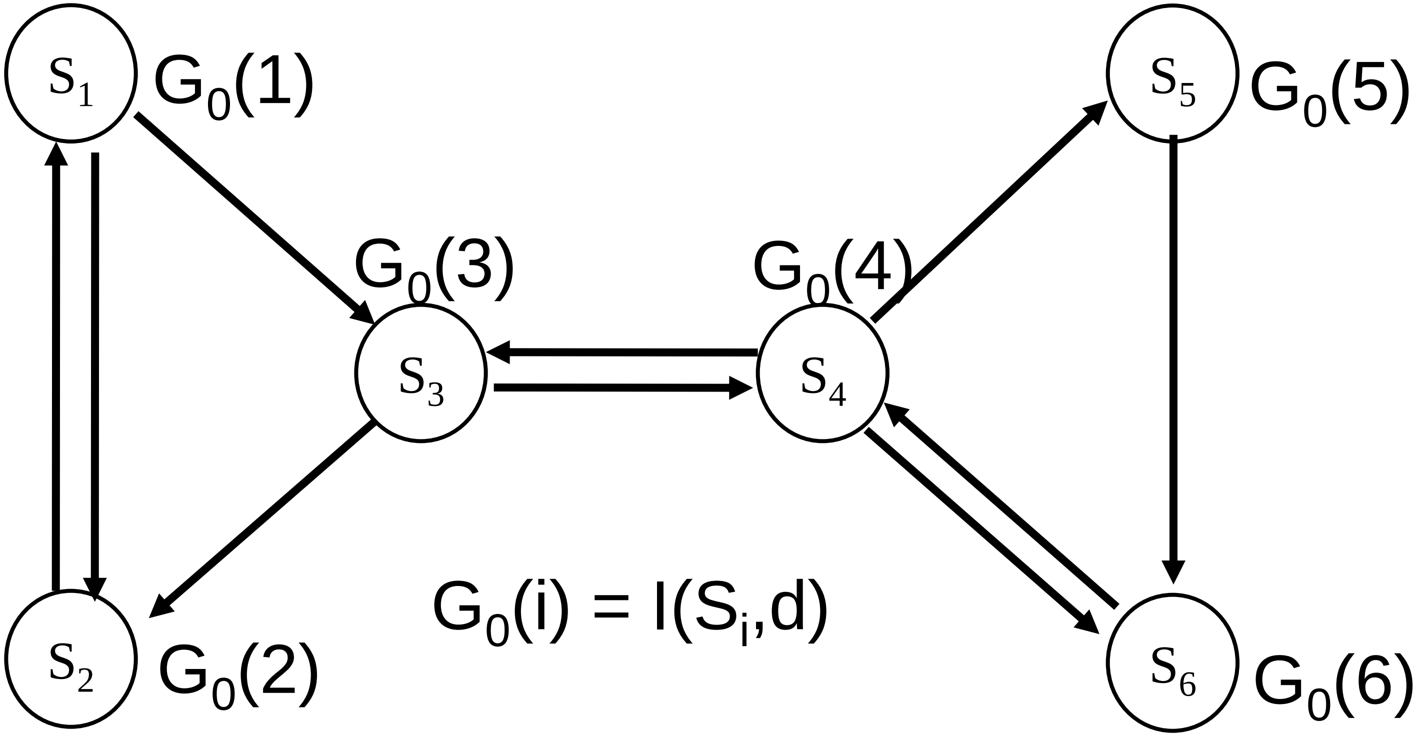
$$G(5) = (1-\alpha)\mathrm{I}(S_5, d) + \alpha G(4)\hat{W}(4,5)$$

$$G(6) = (1-\alpha)\mathrm{I}(S_6, d) + \alpha G(4)\hat{W}(4,6) + \alpha G(5)\hat{W}(5,6)$$

- *How to solve these equations to obtain G(1), G(2), …… G(6)?*
  - *solve the problem iteratively (random walk)*

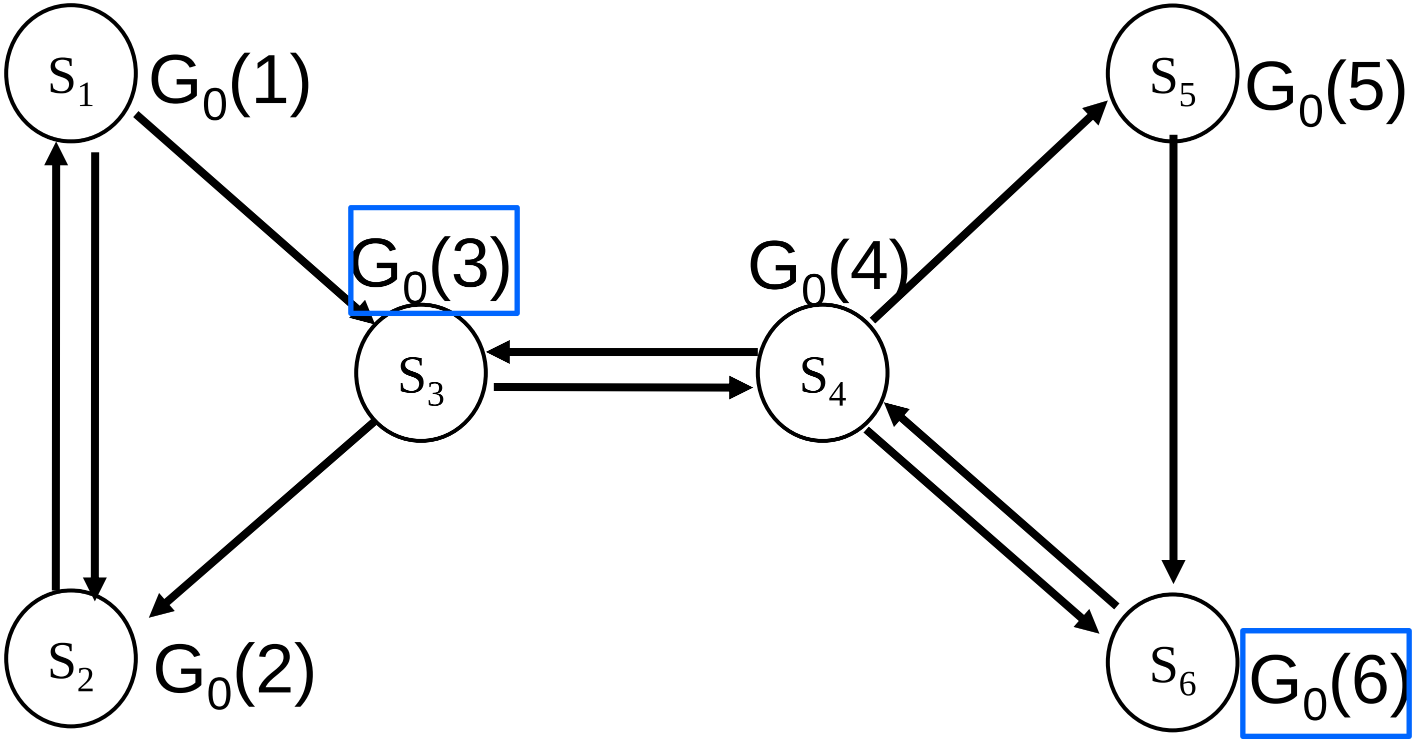
Random Walk Solution

- *Each sentence is assigned an initial value $G_0(i)$*



Random Walk Solution

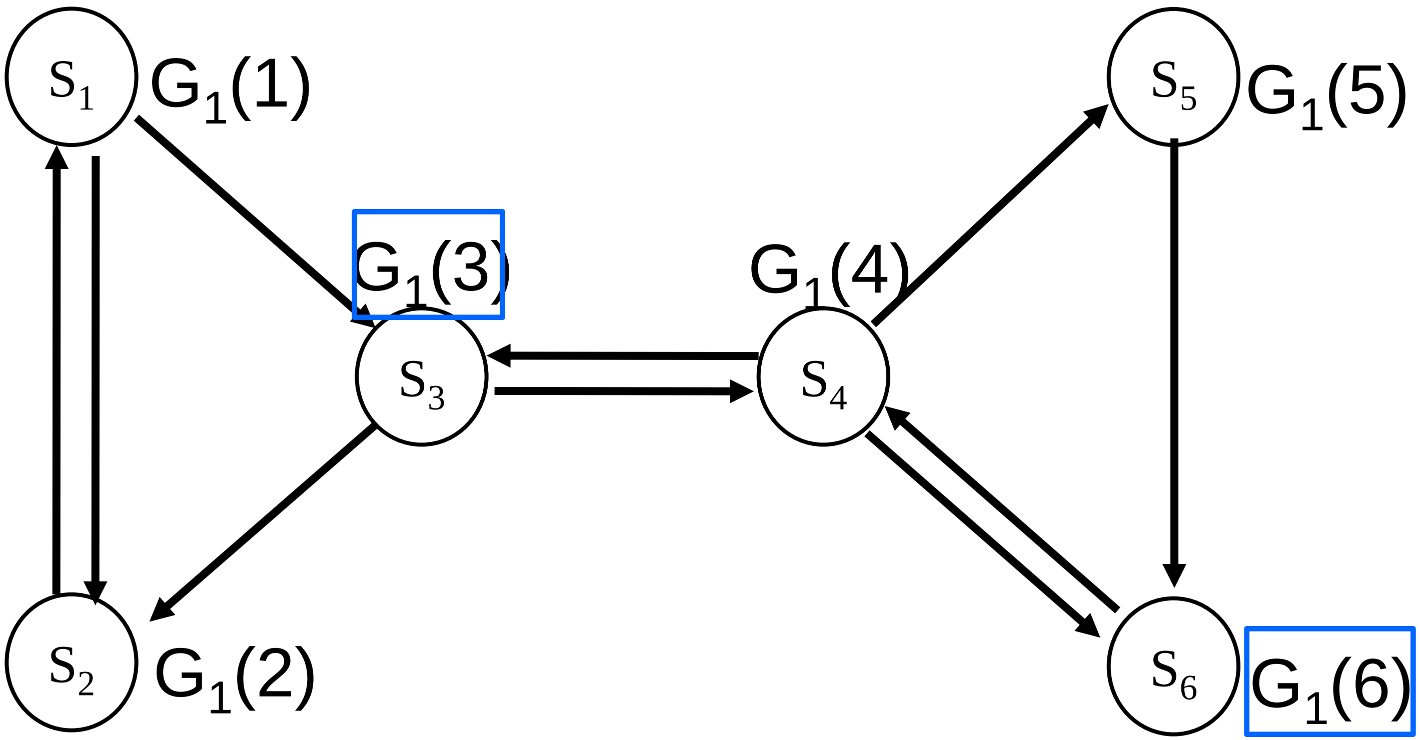
- *Update the score for each sentence ……*



$$\boxed{G_1(4)} = (1-\alpha)\boxed{I(S_4, d)} + \alpha \boxed{G_0(6)} \hat{W}(6,4) + \alpha \boxed{G_0(3)} \hat{W}(3,4)$$

Random Walk Solution

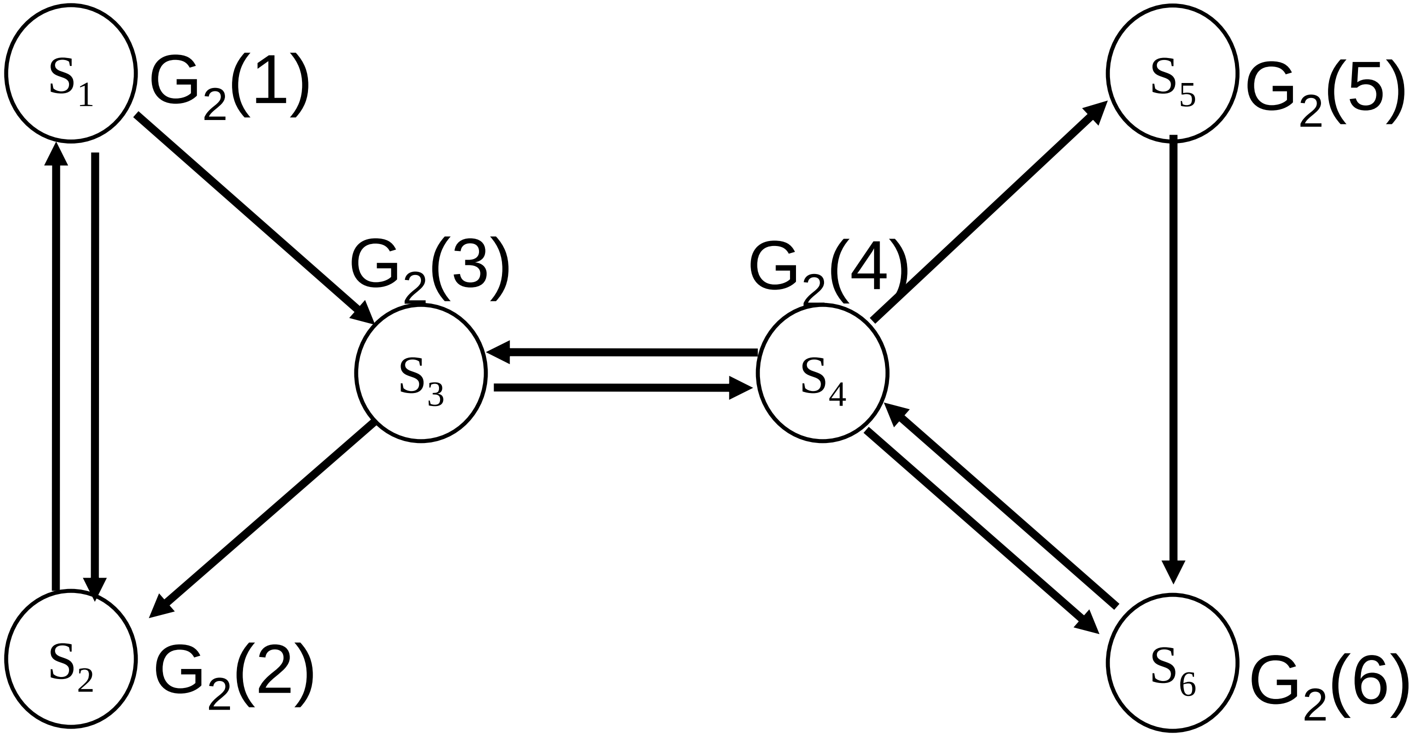
- *Update the score for each sentence ……*



$$\boxed{G_2(4)} = (1-\alpha)\boxed{I(S_4, d)} + \alpha \boxed{G_1(6)} \hat{W}(6,4) + \alpha \boxed{G_1(3)} \hat{W}(3,4)$$

Random Walk Solution

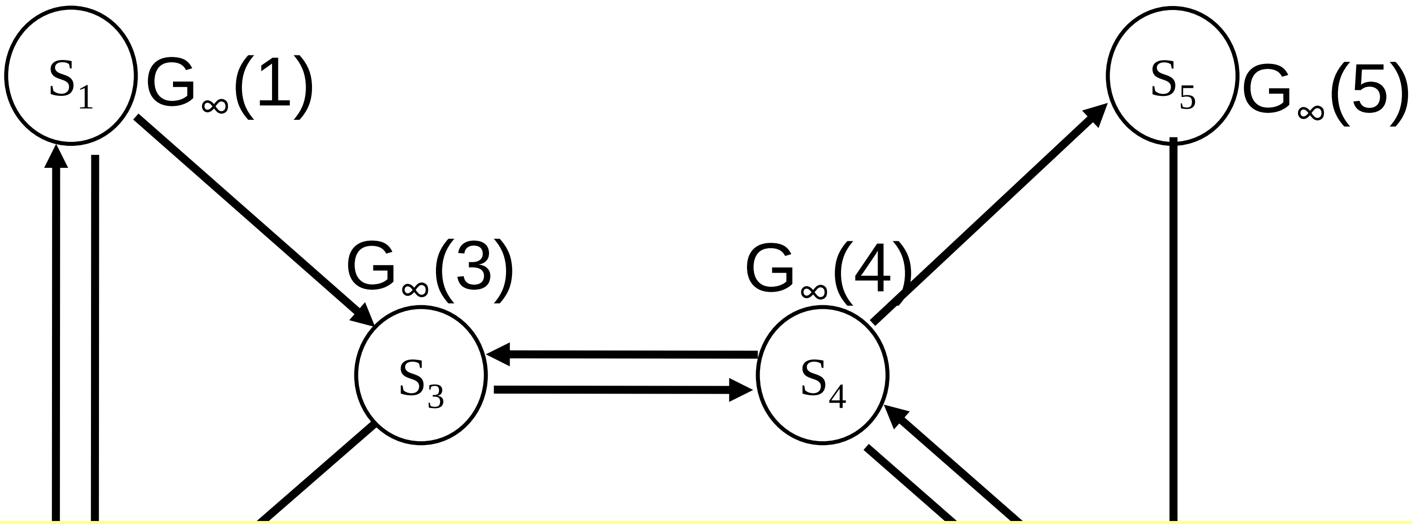
◆ *The process is repeated ……*



$$G_3(4) = (1-\alpha)I(S_4, d) + \alpha G_2(6)\hat{W}(6,4) + \alpha G_2(3)\hat{W}(3,4)$$

Random Walk Solution

- *The process is repeated ……*



- *According to the theory of random walk:*
  - *The score of each node would finally converge.*
- *The converged score $G_\infty(i)$ is actually $G(i)$ satisfying*

$$G(i) = (1-\alpha)\mathrm{I}(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j)\hat{W}(j,i)$$

Graph-based Summarization Approach

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \hat{W}(j, i)$$

$$I(S, d) = \sum_{i=1}^{n} [s(t_i, d) \cdots\cdots] + \cdots\cdots$$

Importance score

term statistical measure

$$\boxed{I'(S_i,d)} = \boxed{I(S_i,d)}^{1-\delta} \boxed{G(i)}^{\delta}$$

New scores: Consider graph structure

For summary selection

Original importance score based on terms in the sentences

Experimental Setup (1/2)

- Corpus: course offered in National Taiwan University
  - Mandarin Chinese embedded by English words
  - Single speaker
  - 45.2 hours
- ASR System
  - Bilingual AM with model adaptation [1]
  - LM with adaptation using random forests [2]

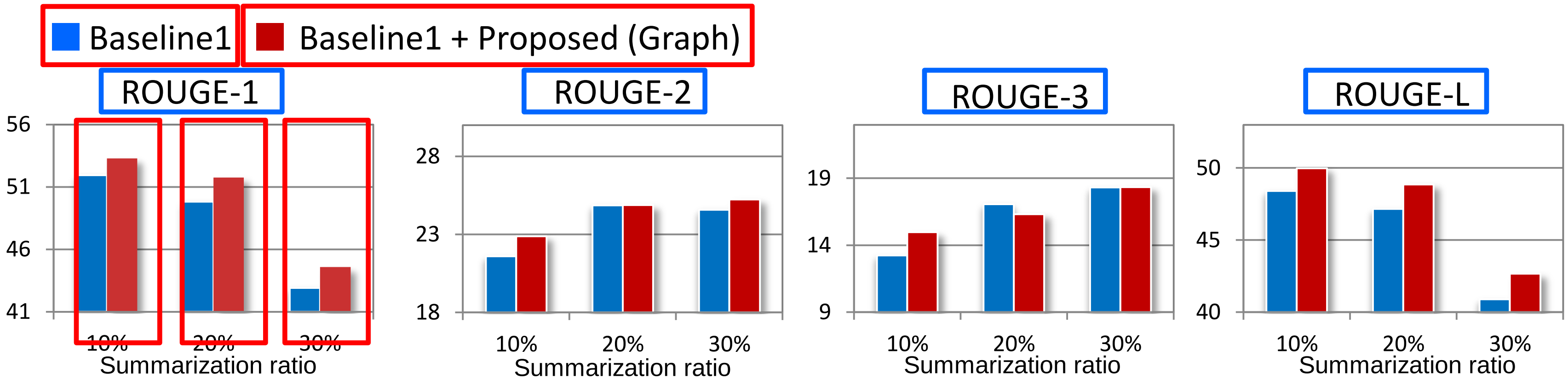
Language	Mandarin	English	Overall
Acc (%)	78.15	53.44	76.26

[1] Ching-Feng Yeh, et al., “Bilingual Acoustic Model Adaptation by Unit Merging on Different Levels and Cross-level Integration, ” Interspeech, 2011.
[2] Ching-Feng Yeh, et al. , “An Integrated Framework for Transcribing Mandarin-English Code-mixed Lectures with Improved Acoustic and Language Modeling,” ISCSLP, 2010.

Experimental Setup (2/2)

- Spoken Documents
  - We segmented the whole lecture into 155 documents by topic segmentation
  - 34 documents out of the 155 were tested.
  - The average length of each document was about 17.5 minutes
  - Human produced reference summaries for each document
- Evaluation
  - ROUGE-1, ROUGE-2, ROUGE-3
  - ROUGE-L: Longest Common Subsequence (LCS)

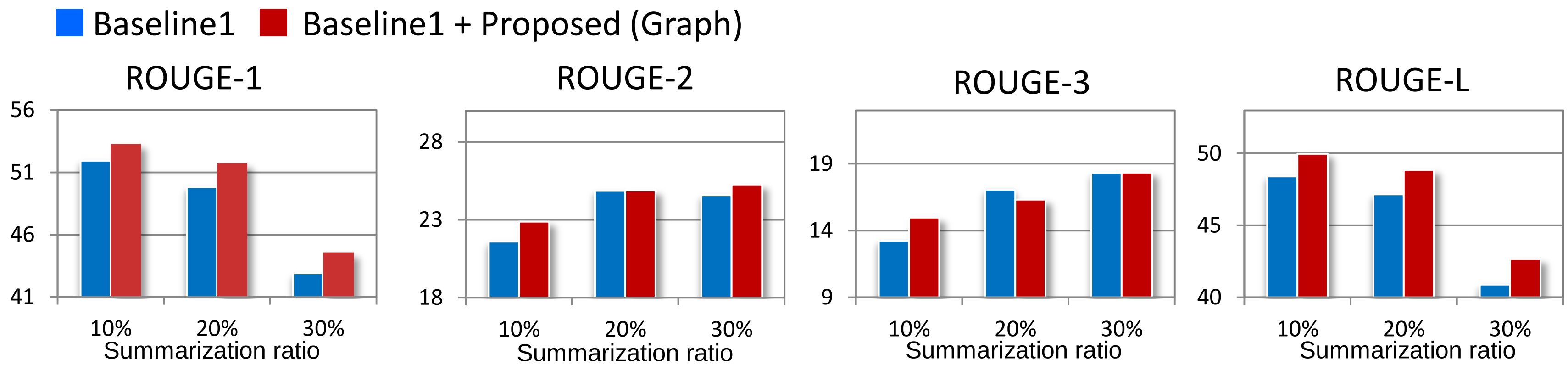
Experimental Results



Baseline1: $I(S_i,d)$ – importance score using latent topic entropy term statistical measure

Baseline1+Proposed: $I(S_i,d)G(i)$

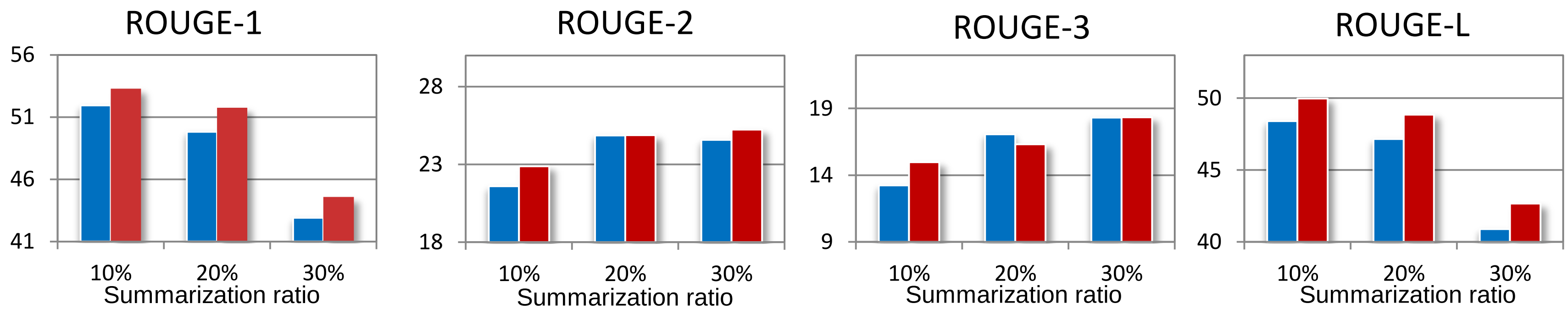
Experimental Results



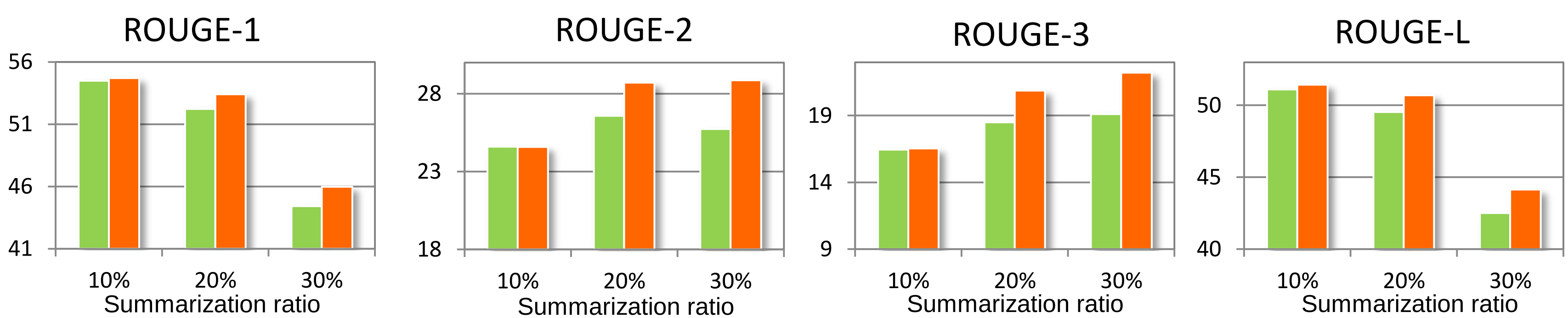
- The proposed approach outperformed the first baseline in most cases. (Compare blue and red bars)

Experimental Results

Baseline1 | Baseline1 + Proposed (Graph)



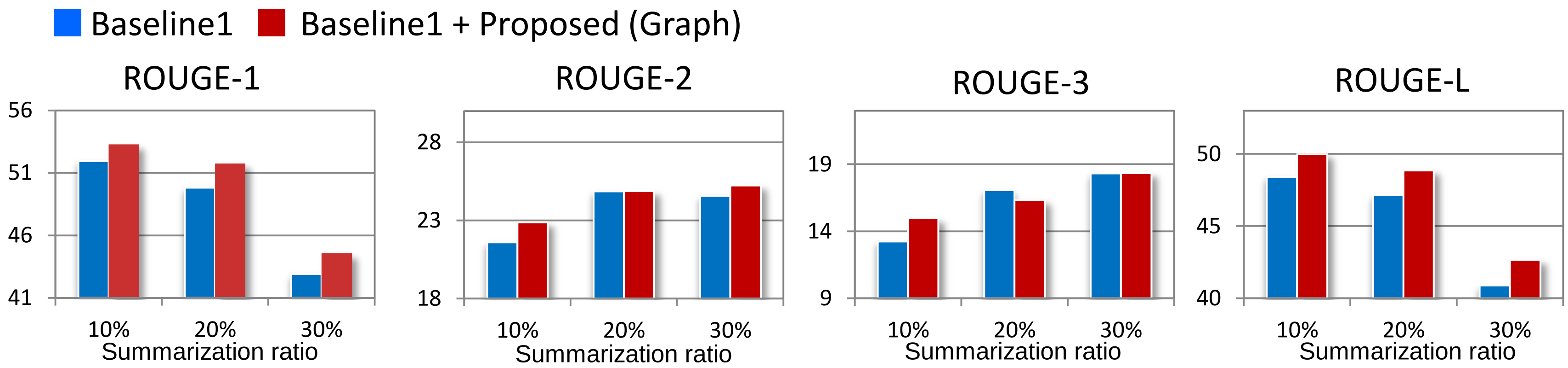
Baseline2 | Baseline2 + Proposed (Graph)

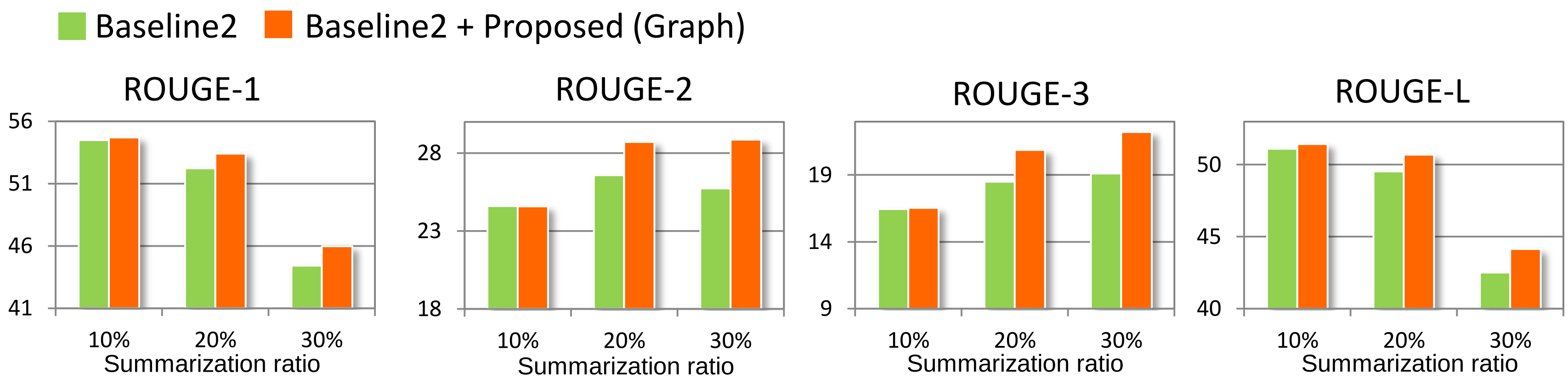


Baseline2: $I(S_i,d)$ – importance score using key-term based statistical measure

Baseline2+Proposed: $I(S_i,d)G(i)$

Experimental Results





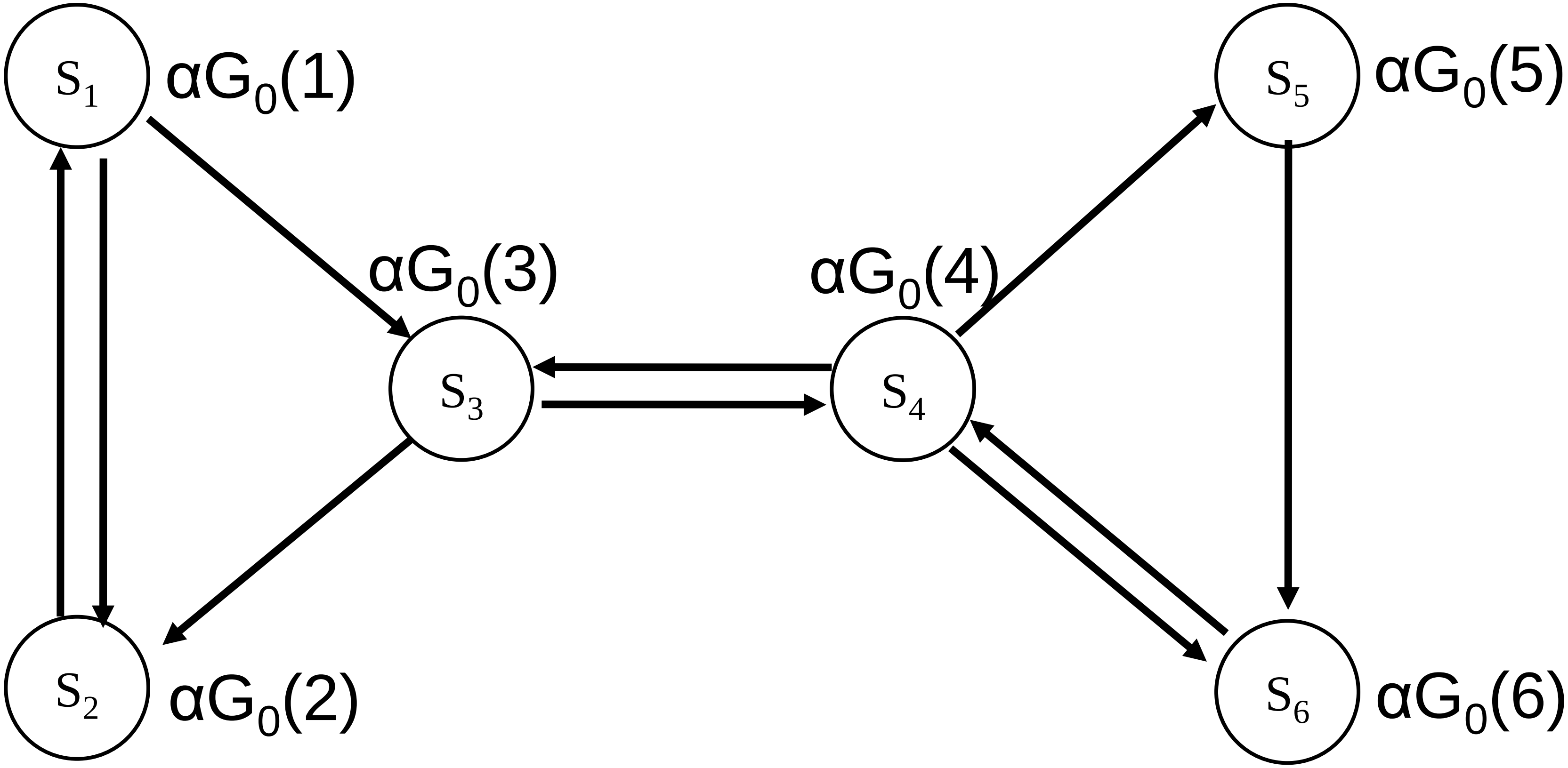
- The proposed approach always outperformed the second baseline. (Compare green and orange bars)

Conclusions

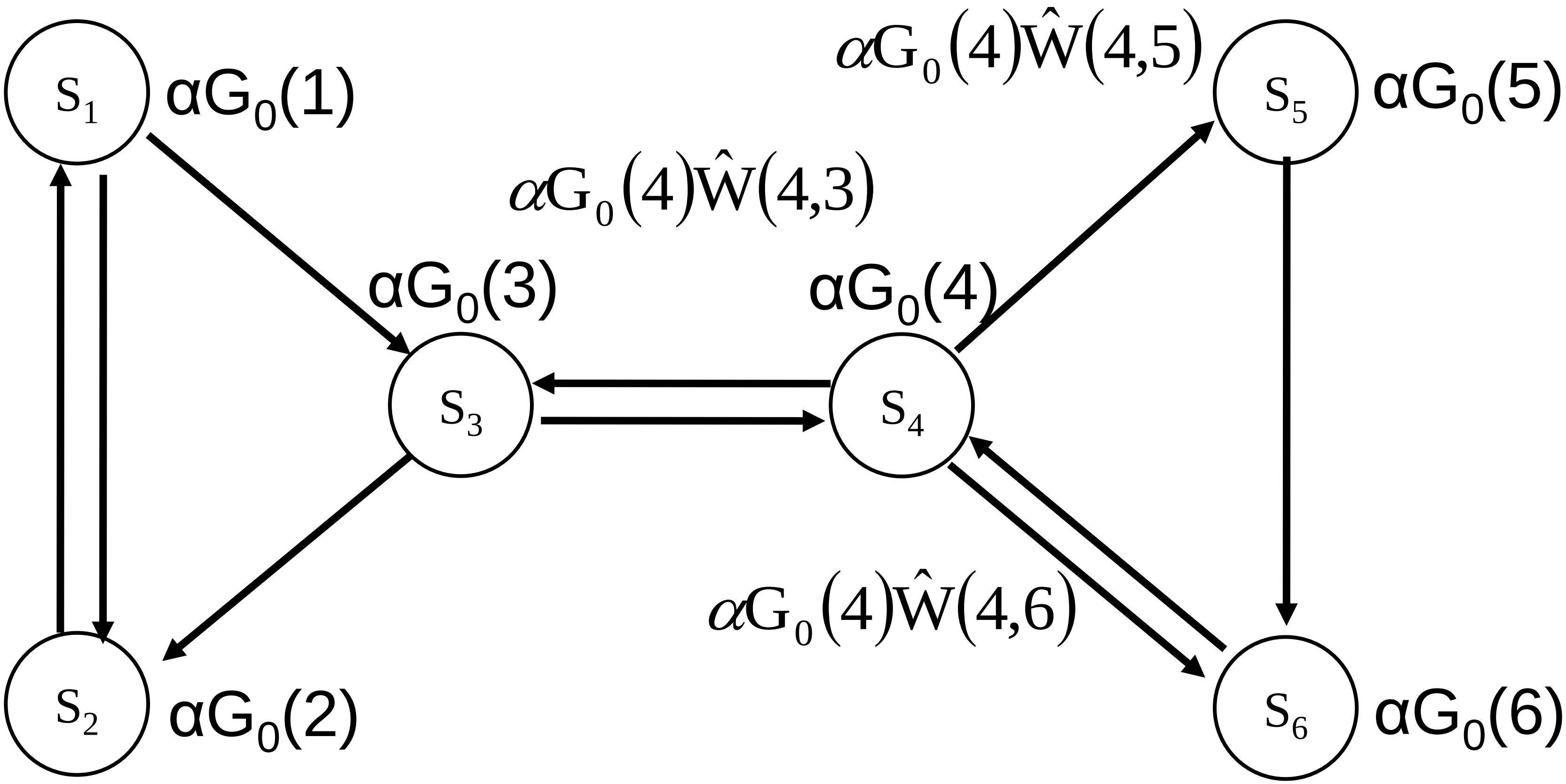
- The performance of summarization can be improved by
  - Graph-based approach considering topical similarity
    - This offers a way to globally consider all sentences in a document for summarization rather than considers each sentence individually

Thanks for your attention! ☺
Q & A

Random Walk Solution

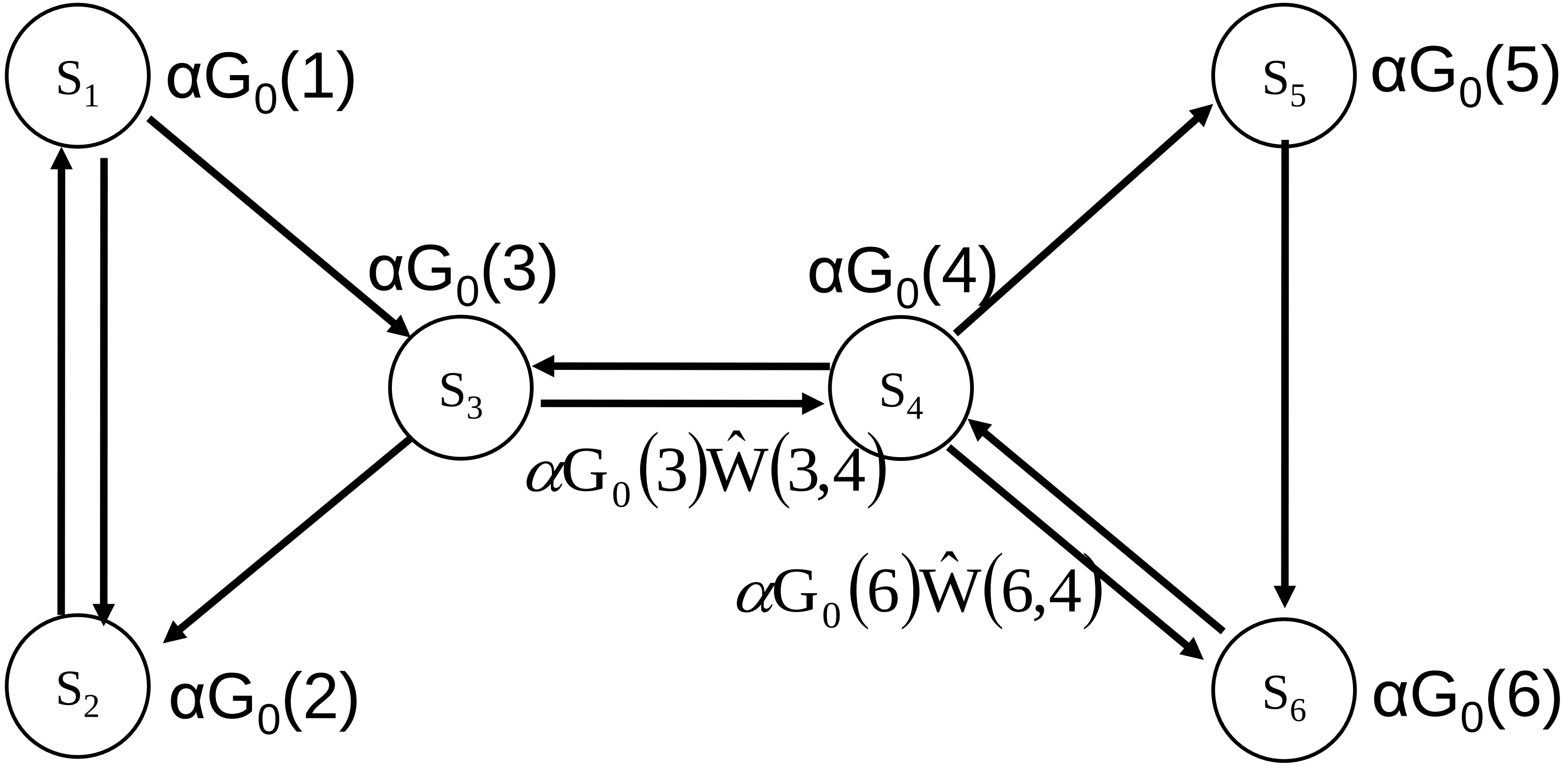


Random Walk Solution

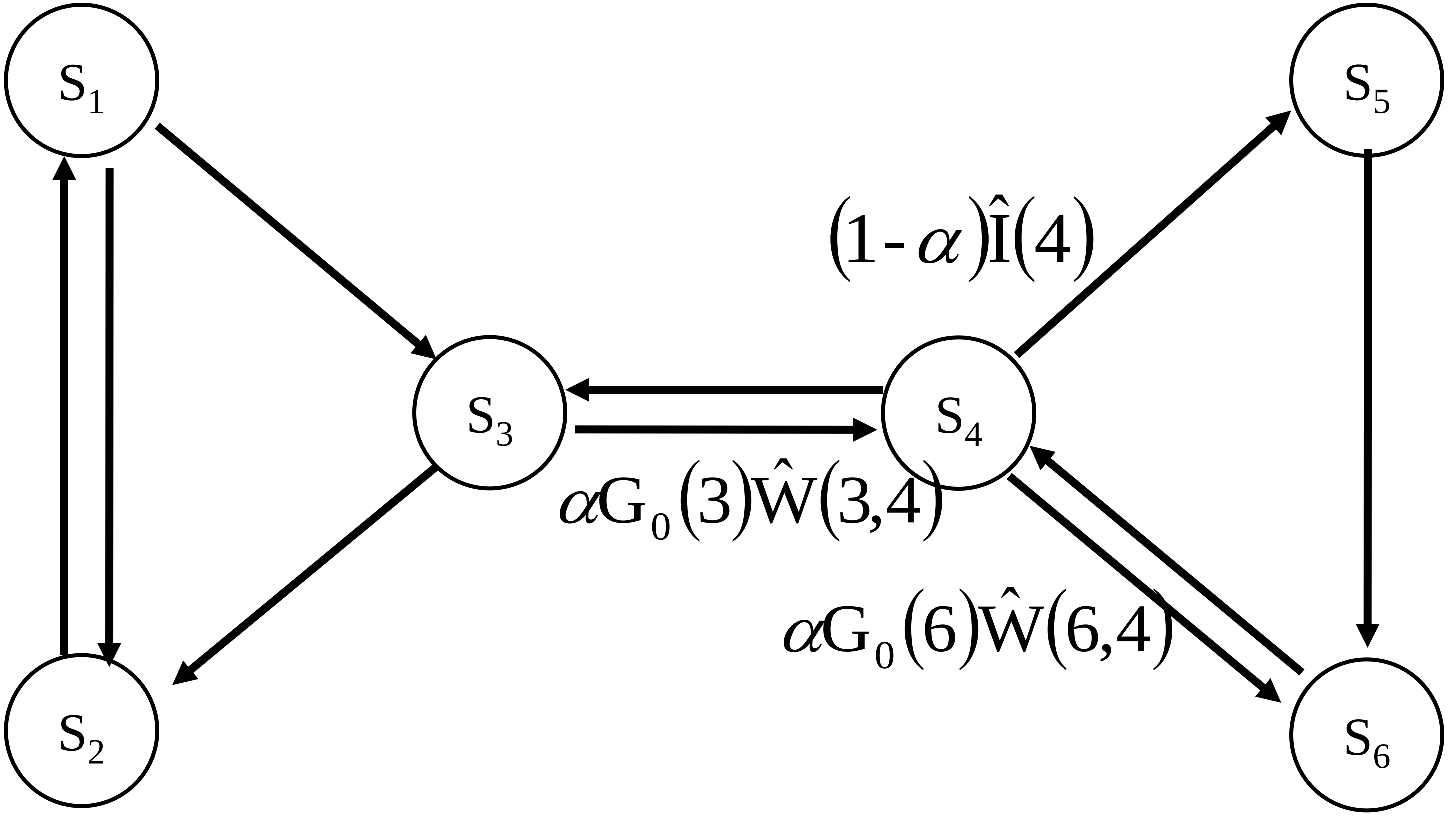


$$\hat{W}(j,i) = \frac{W(j,i)}{\sum_{S_k \in out(j)} W(j,k)}$$

Random Walk Solution



Random Walk Solution



$$G_1(4) = (1-\alpha)\hat{I}(4) + \alpha G_0(3)\hat{W}(3,4) + \alpha G_0(6)\hat{W}(6,4)$$

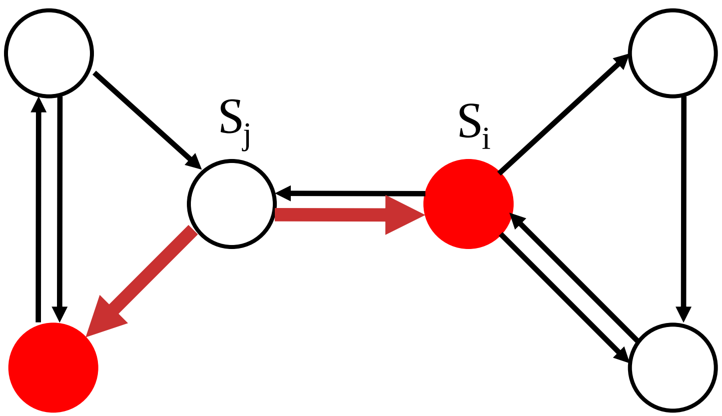
Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha)I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j)\boxed{\hat{W}(j,i)}$$

$$\hat{W}(j,i) = \frac{W(j,i)}{\boxed{\sum_{S_k \in out(j)} W(j,k)}}$$

$$\Rightarrow \sum_{S_k \in out(j)} \hat{W}(j,k) = 1$$



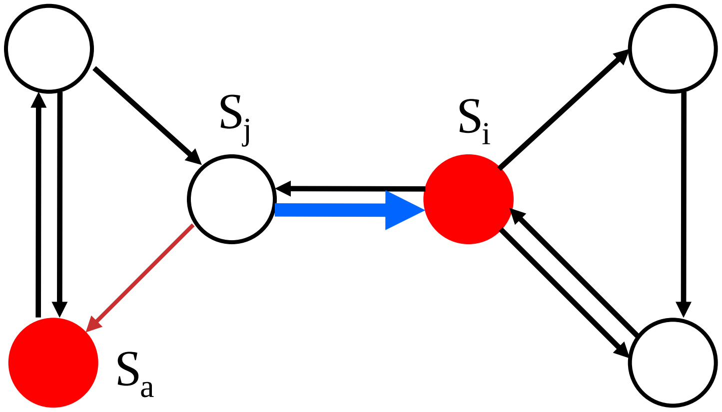
Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \hat{W}(j,i)$$

The amount of score S_j propagate to S_i is

$$G(j) \frac{W(j,i)}{W(j,i) + W(j,a)}$$



Mathematical Formulation

Find a set of new scores based on graph structure {G(i) for each sentence S_i in document d} which satisfies

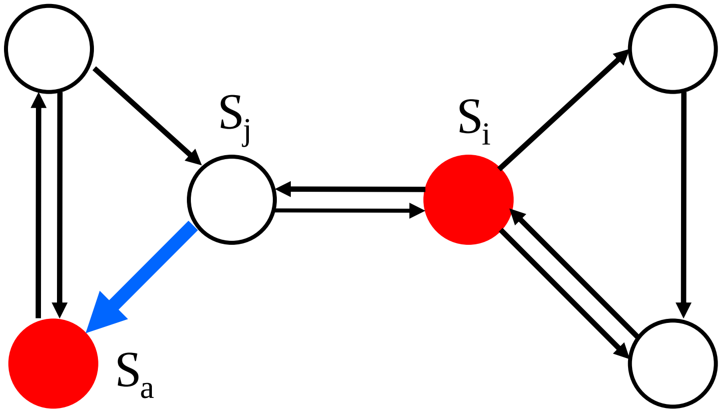
$$G(i) = (1-\alpha) I(S_i, d) + \alpha \sum_{S_j \in in(i)} G(j) \hat{W}(j,i)$$

The amount of score S_j propagate to S_i is

$$G(j) \frac{W(j,i)}{W(j,i) + W(j,a)}$$

The amount of score S_j propagate to S_a is

$$G(j) \frac{W(j,a)}{W(j,i) + W(j,a)}$$

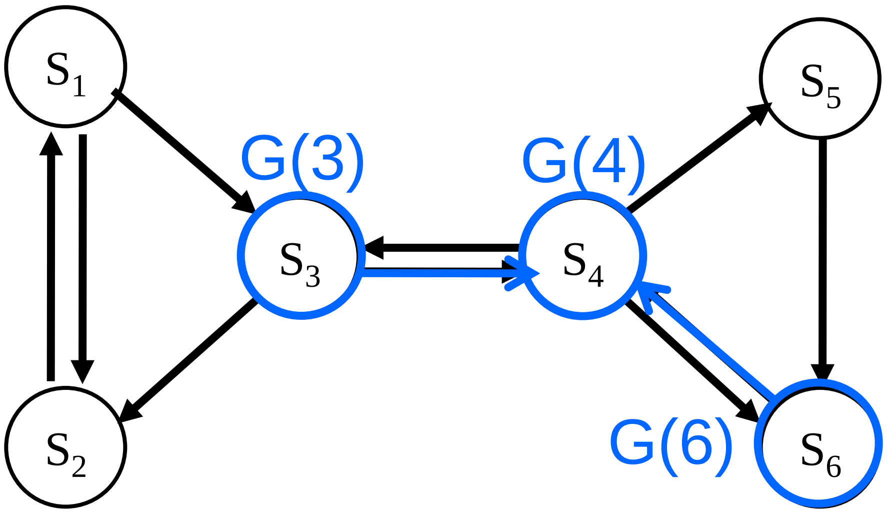


Mathematical Formulation – an Example

Find G(1), G(2), G(3), G(4), G(5), G(6) such that

$$\hat{W}(3,4) = \frac{W(3,4)}{W(3,2)+W(3,4)}$$

$$\hat{W}(6,4) = W(6,4)$$



$$\boxed{G(4)} = \underbrace{(1-\alpha)\boxed{I(S_4, d)}}_{\text{depends on } S_4 \text{ itself}} + \underbrace{\alpha \boxed{G(6)}\boxed{\hat{W}(6,4)} + \alpha \boxed{G(3)}\boxed{\hat{W}(3,4)}}_{\text{Depends on topically similar sentences } (S_3 \text{ and } S_6)}$$