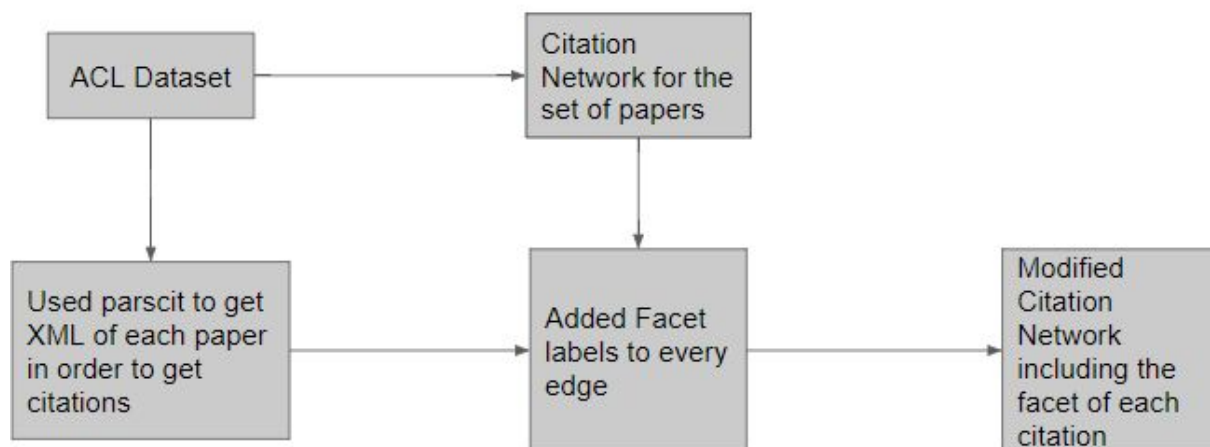


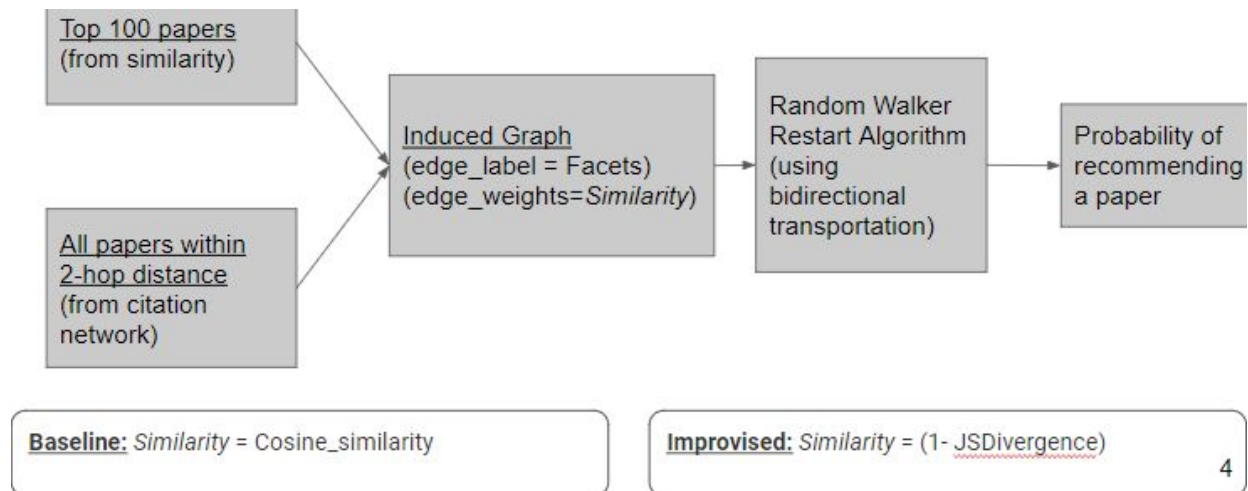
# Designing Faceted Recommendation System for Scientific Articles

## Introduction

- A smart recommendation engine should be able to organize the recommended papers into multiple facets/tags such as background, alternative approaches, methods and comparison.
- We have used the AAN dataset which is an assemblage of all papers included in ACL2 publication venue and categorize the citation links based on their occurrence in various sections of the paper.

## Flowchart of Experiments





## Random Walk with Restarts

- Induced Citation Graph: From the citation network, a directed induced graph corresponding to the query paper is generated.
- Baseline: (**hop-limit = 2 or cosine\_similarity > 0.25**)  
Improved: (**hop-limit = 2 or JSDiv < 0.51**)
- Weighted edges: Each edge is assigned a probability based on the (**similarity**) or (**1-divergence**) of corresponding nodes(papers).

$$p^{t+1} = (1-r) A p^t + p^0$$

A: Transportation Matrix  
 P<sup>t</sup>: probability matrix at time = t  
 r : restart probability = 0.4

- The relevance score of node  $j$  with respect to node  $i$  is defined by the steady-state probability  $r_{ij}$  that the walker will finally stay at node  $j$ .

## Improvisation of the model using Latent Dirichlet Allocation

- Topic Modeling using Latent Dirichlet Allocation(LDA).
- Num\_topics = 100

**Jensen–Shannon Divergence:** Method for measuring the similarity between two probability distributions.

$$JSD(P \parallel Q) = \frac{1}{2}D(P \parallel M) + \frac{1}{2}D(Q \parallel M)$$

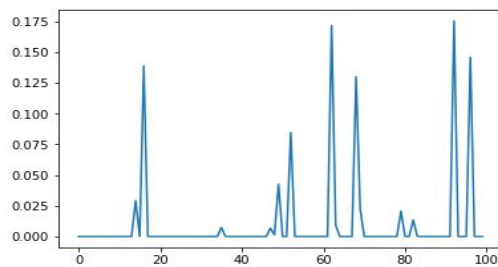
$$M = \frac{1}{2}(P + Q)$$

**Results:**

Topic Probability distribution for 'A00-1009'

```
plt.plot(get_doc_topics(ldamodel, corpus[paper_array.index('A00-1009')]))
```

Out[42]: [



**A00-2004 (Citations for Method Facet)**

**Advances In Domain Independent Linear Text Segmentation**

**Cosine Similarity**

J06-3003

Similarity of Semantic Relations

C10-1142

Estimating Linear Models for Compositional Distributional Semantics

P93-1001

Char Align: A Program For Aligning Parallel Texts At The Character Level

N01-1027

Identifying User Corrections Automatically In Spoken Dialogue Systems

**Jensen Shannon Divergence**

P93-1001

Char Align: A Program For Aligning Parallel Texts At The Character Level

W01-0514  
Latent Semantic Analysis For Text Segmentation

A94-1013  
Adaptive Sentence Boundary Disambiguation

A97-1004  
A Maximum Entropy Approach To Identifying Sentence Boundaries

## Evaluation

Surveying with 4 people in our group, **JSDivergence** performed better than **Cosine\_similarity** 60% of the times.

Name	Subject-1	Subject-2	Subject-3	Subject-4
Set-1 (Cosine Similarity)	4	5	4	3
Set-2 (JSDivergence)	6	5	6	7