Random Walks Analysis on Graph Modelled Multimodal Collections



Serwah Sabetghadam
Supervisors: Andreas Rauber, Mihai Lupu
Technical University of Vienna, IFS group

@serwah





In addition to Multimodal domain specific collections







Multimodal Retrieval has become a challenge

Multimodal Retrieval

- Many papers on
  - Image search, music search
  - Tag based video search
  - Content based Image retrieval
  - Multimodal reranking (Mei et al. 2014)



- **Info. objects are not isolate**

Mei T., Rui Y., Li S., and Tian Q., Multimedia Search Reranking: A Literature Survey. ACM Computing Surveys, Sept. 2014

Multimodal Data is Interlinked



Structured data represented by graphs

Astera: a Model for Multimodal IR

- Key features
  - Model as a graph
  - Define different relations
    - Similarity
    - Semantic
    - Part-of
    - Facet







S. Sabetghadam, M. Lupu, and A. Rauber, "A combined approach of structured and non-structured IR in multimodal domain," in *ICMR*, 2014.

Traversal Methods in a Graph

- Find relevant documents for a specific query
- Different approaches for graph traversal
- One well-known method:

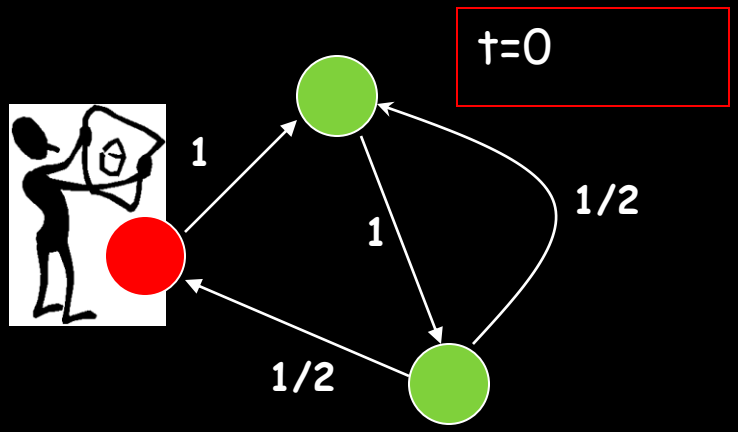
Random Walks (RW)

Challenges

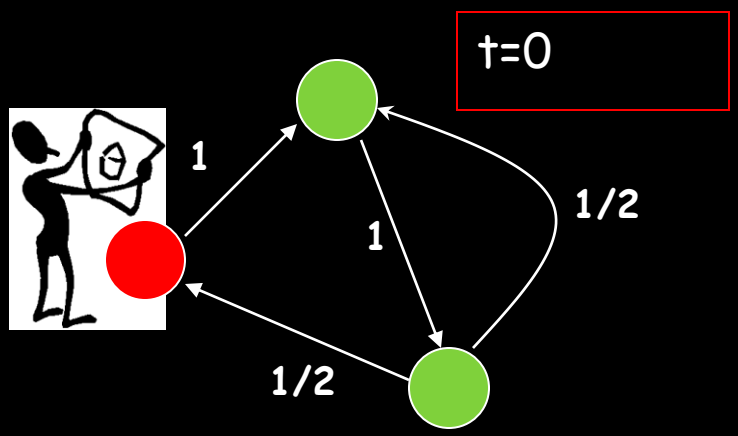
- What is the effect of Query-dependent and Query-independent RandomWalks on our Multimodal graph (Astera)

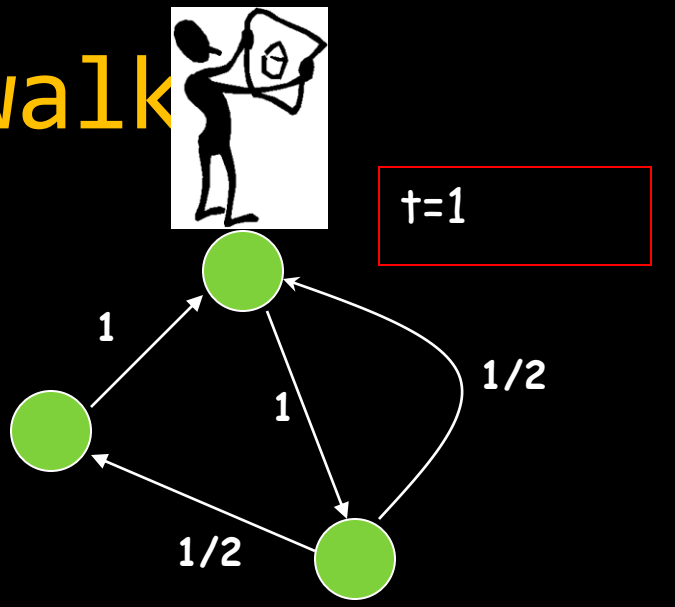
- How they differ in final performance in the graph?

What is a random walk

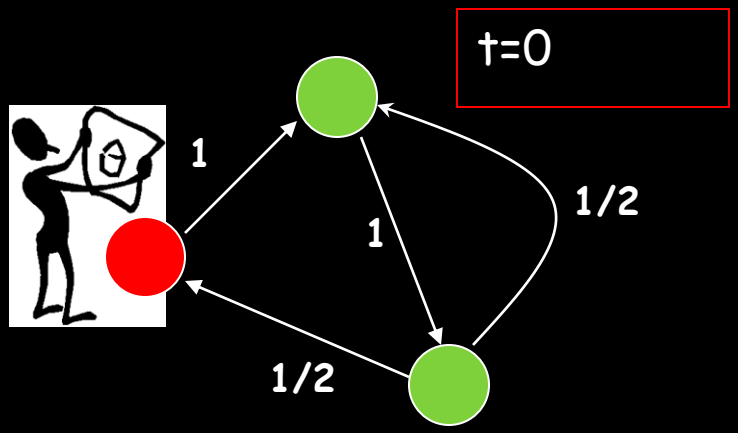


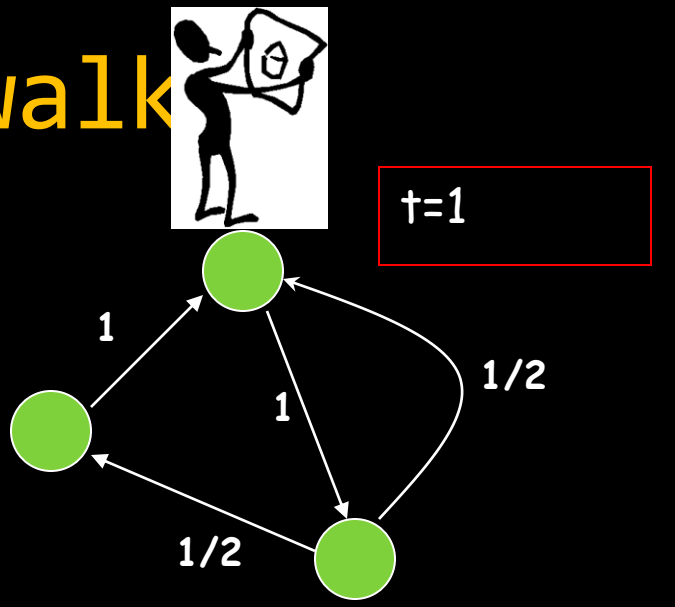
What is a random walk

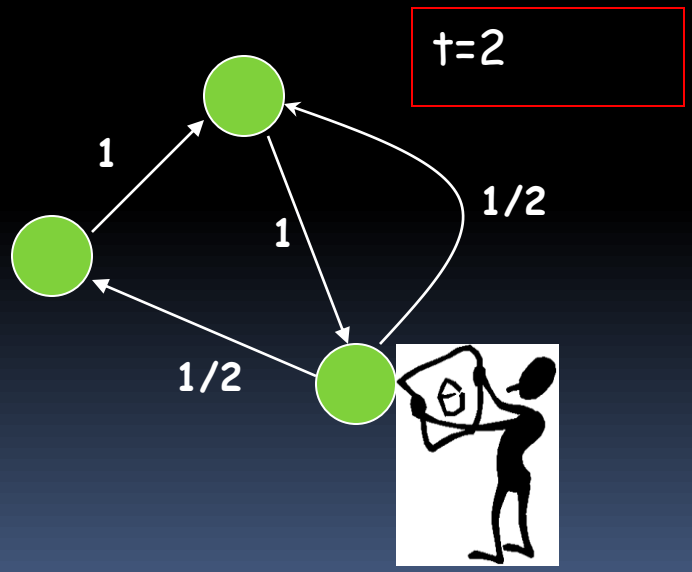




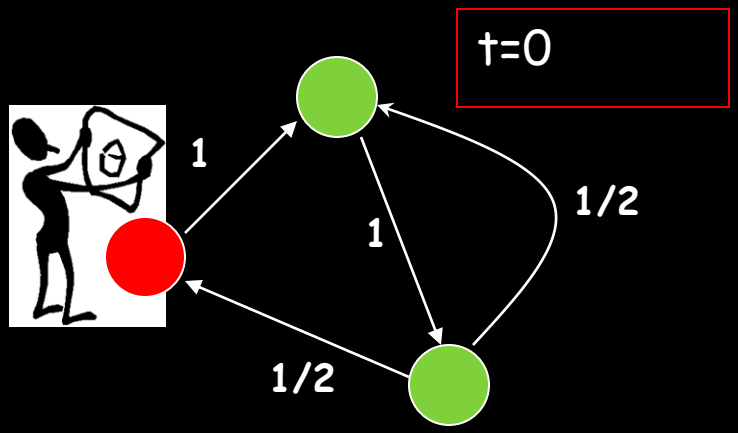
What is a random walk

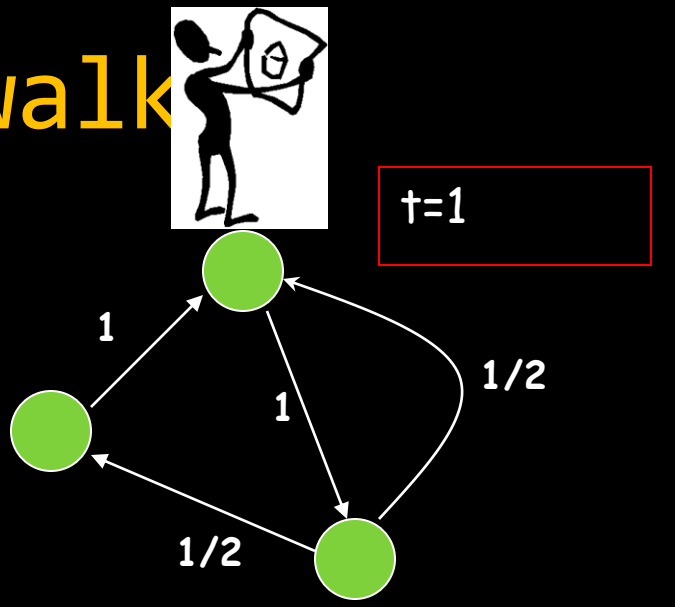


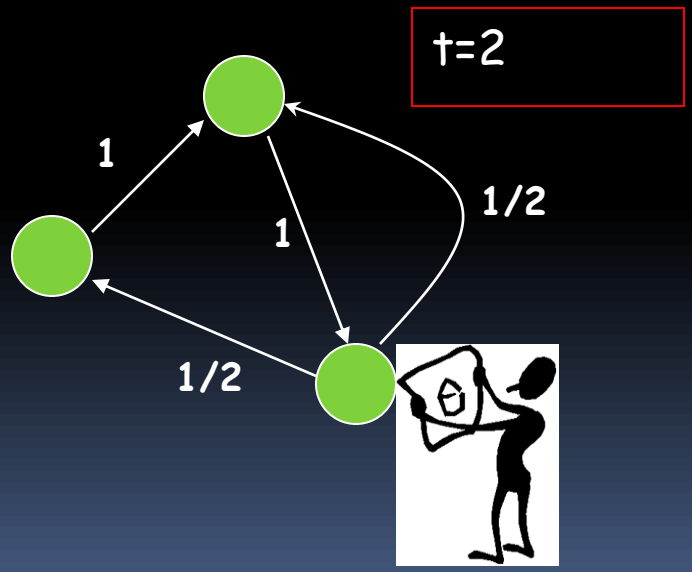


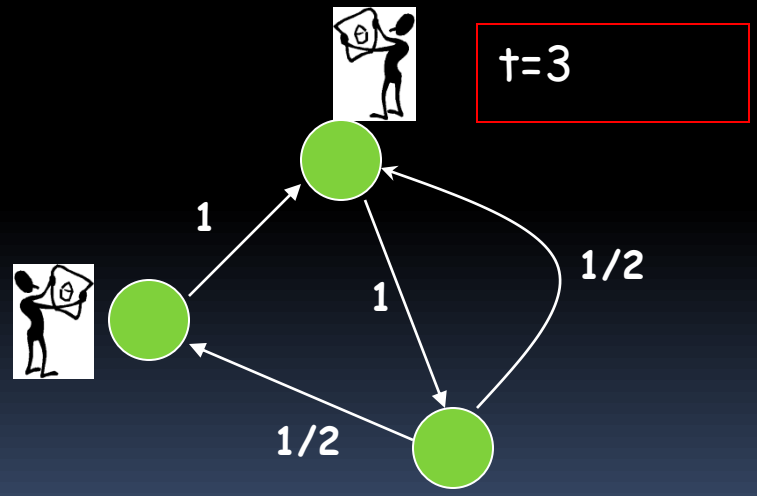


What is a random walk





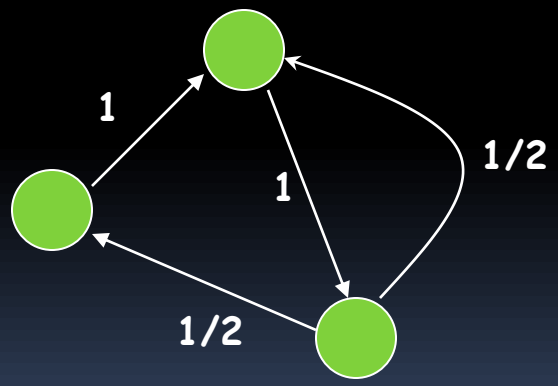




Probabilities in a Transition Matrix

0	1	0
0	0	1
1/2	1/2	0

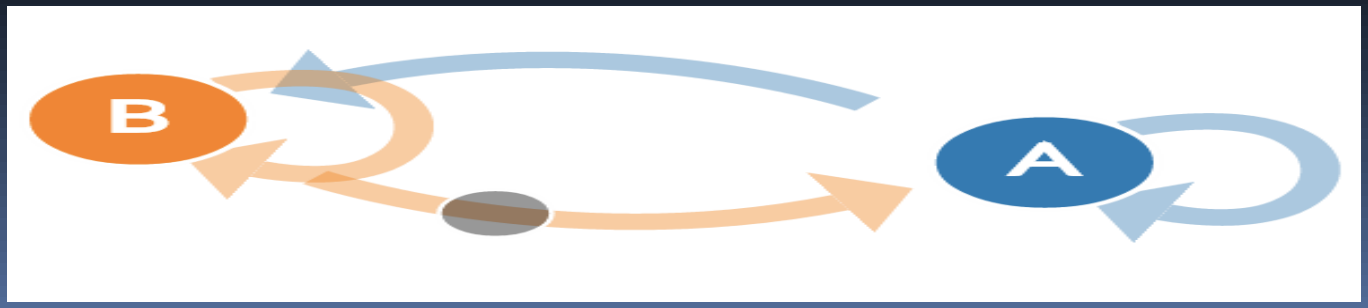
Transition matrix P



Stationary Distribution

- When the surfer keeps walking for a long time
- When the distribution does not change anymore
  - i.e. $x_{t+1} = x_t$

Equilibrium probability distribution = stationary distribution



Challenges

- What is the effect of Query-dependent and Query-independent RandomWalks on our Multimodal graph (Astera)

- How they differ in final performance in the graph?

Query-independent Random Walks

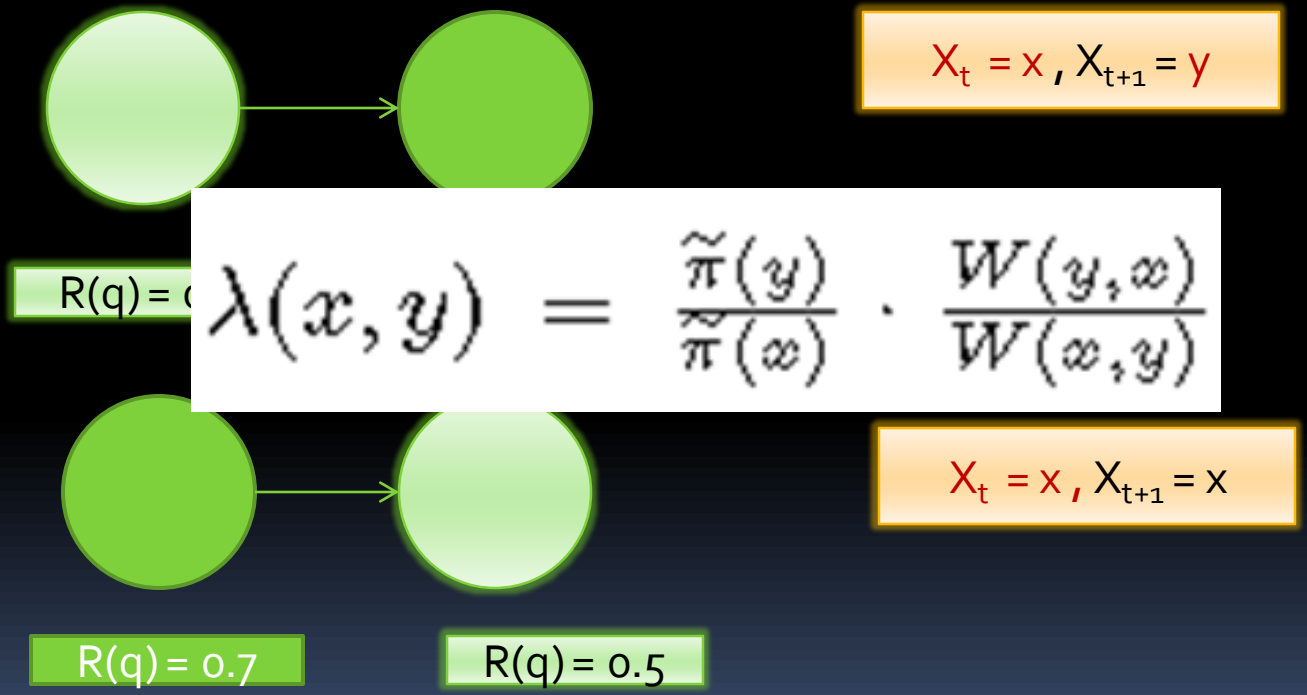
- The basic definition of RW does not consider relevancy to the query
- What we need is only the transition matrix

Query-dependent RW

- We use Metropolis-Hastings (MH) algorithm as query-dependent Random Walks

- In each step
  - Relevancy of source and target nodes to the query are considered

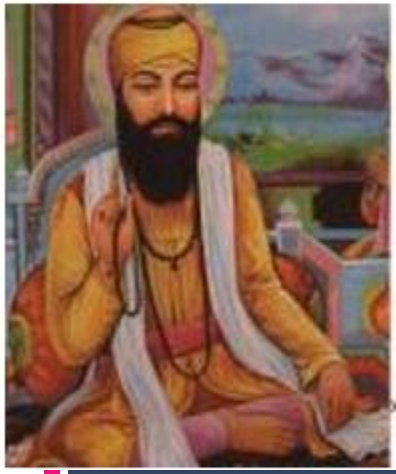
Query dependent RW

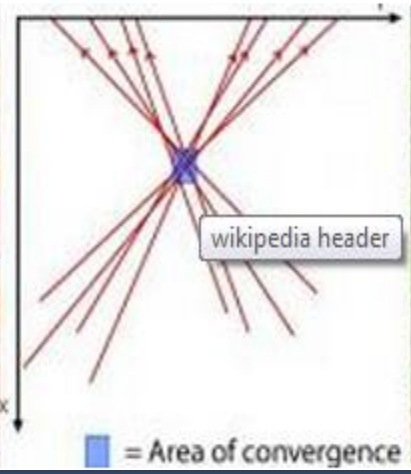


Data collection



- ImageCLEF 2011 Wikipedia collection
- About 400,000 Docs and Imgs
  - 125,828 doc, 273,434 img







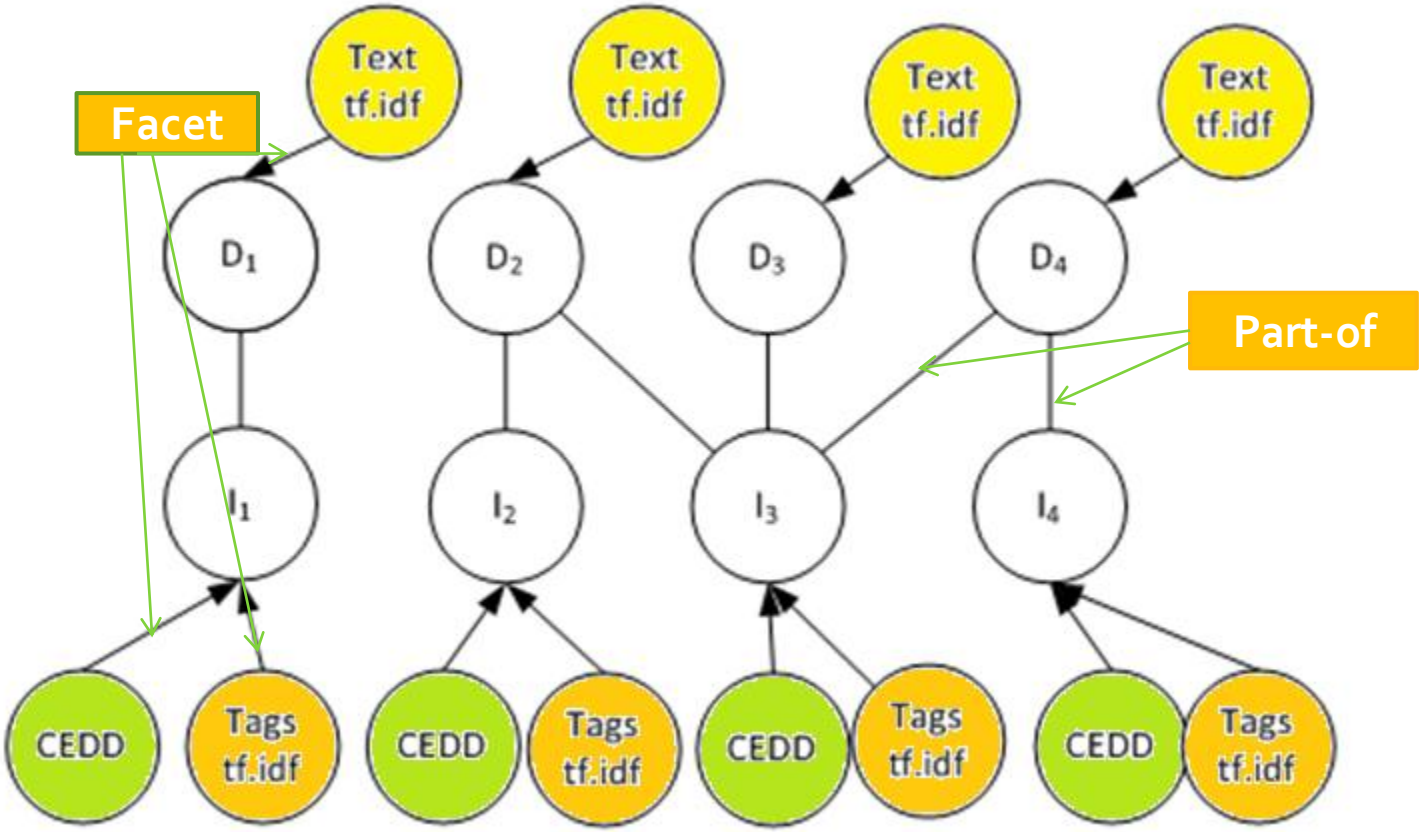




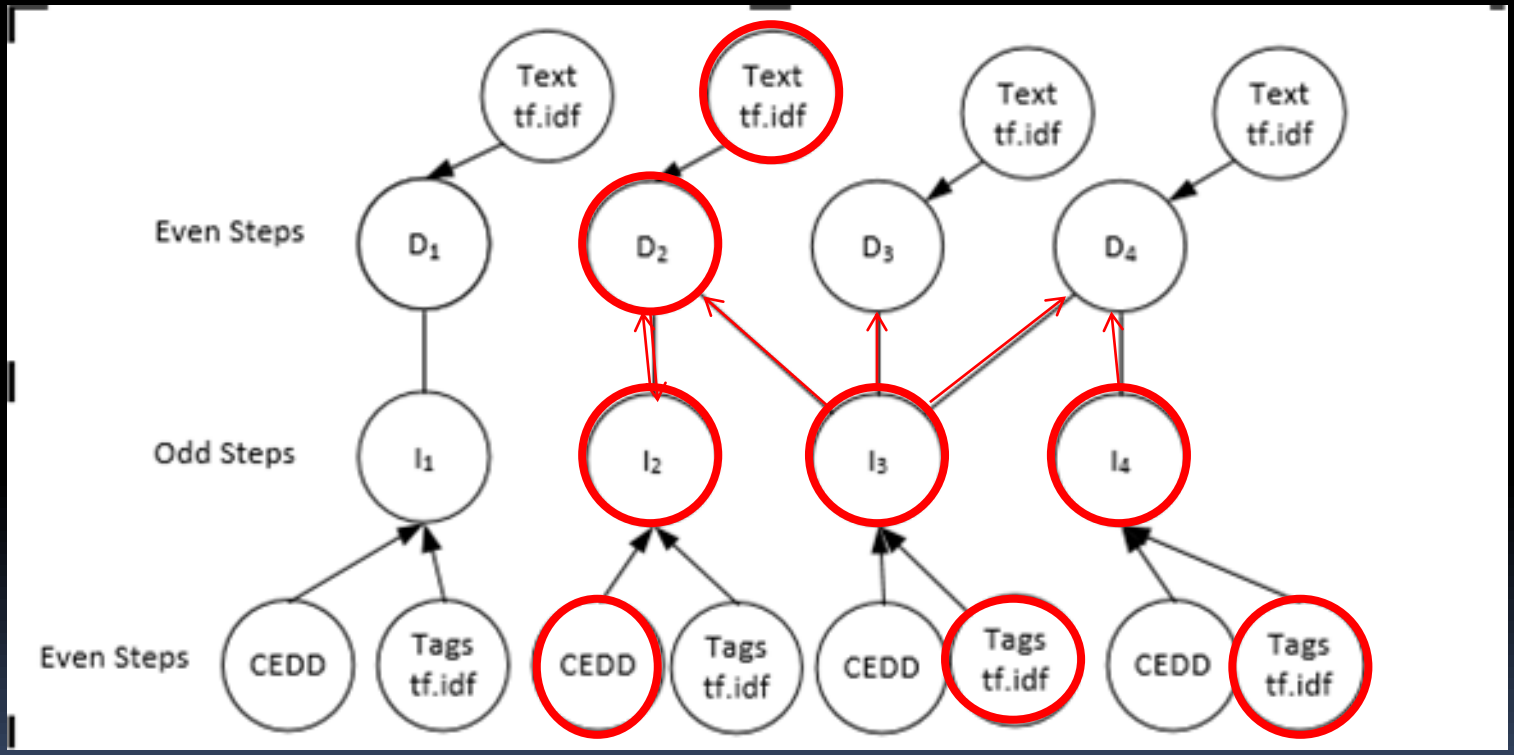
Different indexings based on different features

- Document TF.IDF
- Image tags TF.IDF
- Image CEDD (Color and Edge Directivity Descriptor)

ImageCLEF 2011 Wikipedia modelled

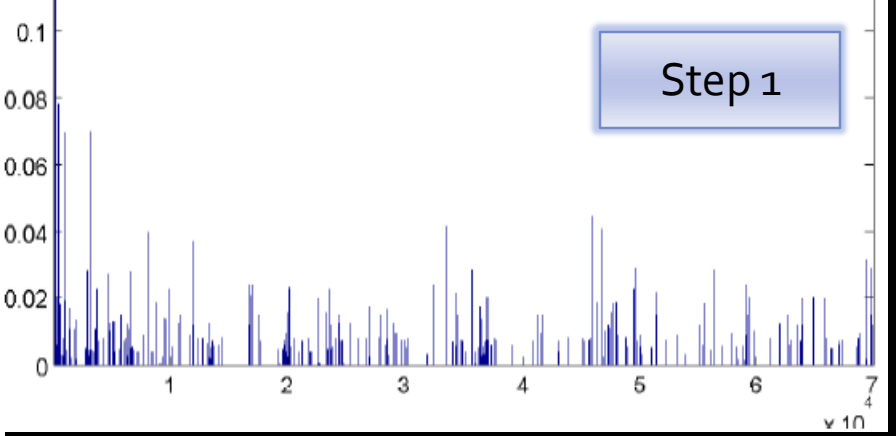


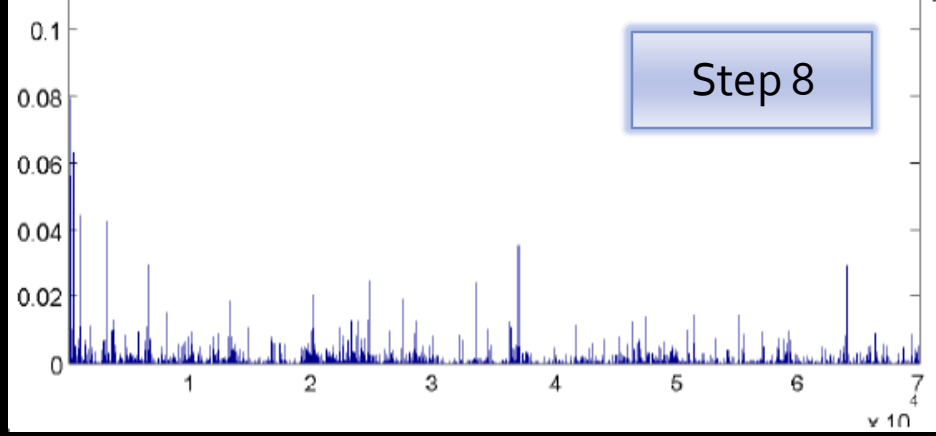
All facets

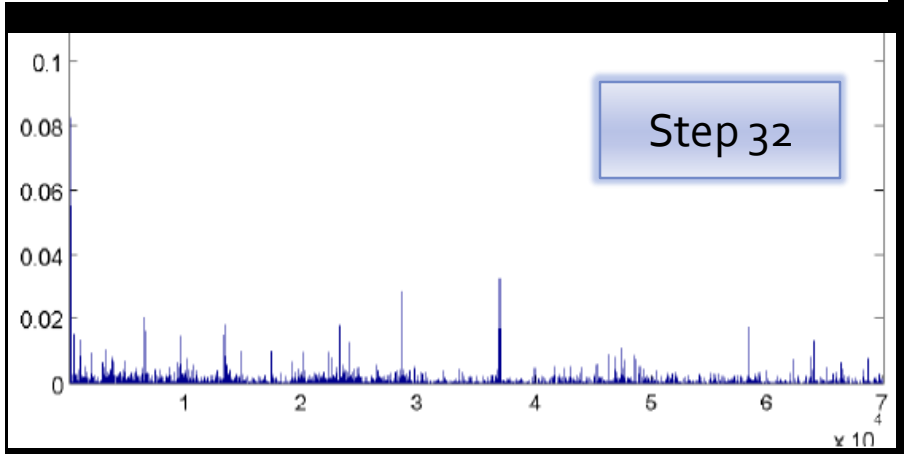

Text tf.idf
Text tf.idf
Text tf.idf
Text tf.idf
Even Steps
D1
D2
D3
D4
Odd Steps
I1
I2
I3
I4
Even Steps
CEDD
Tags tf.idf
CEDD
Tags tf.idf
CEDD
Tags tf.idf
CEDD
Tags tf.idf

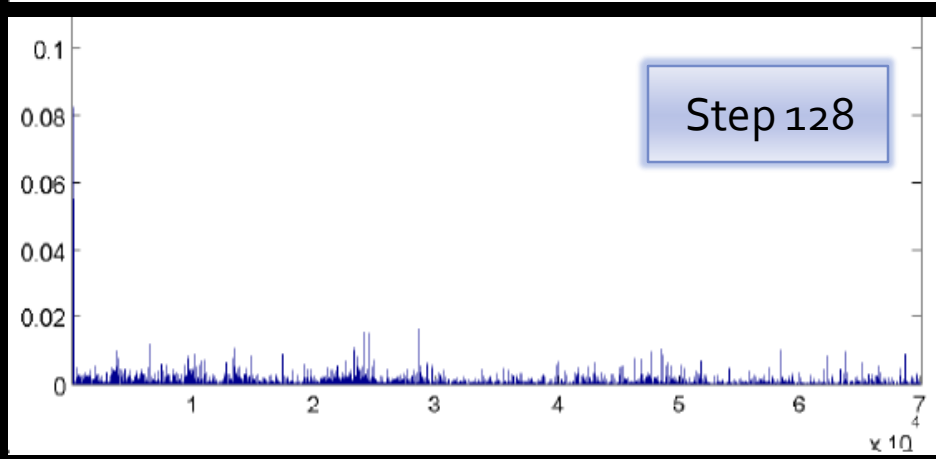
An understanding of score distribution

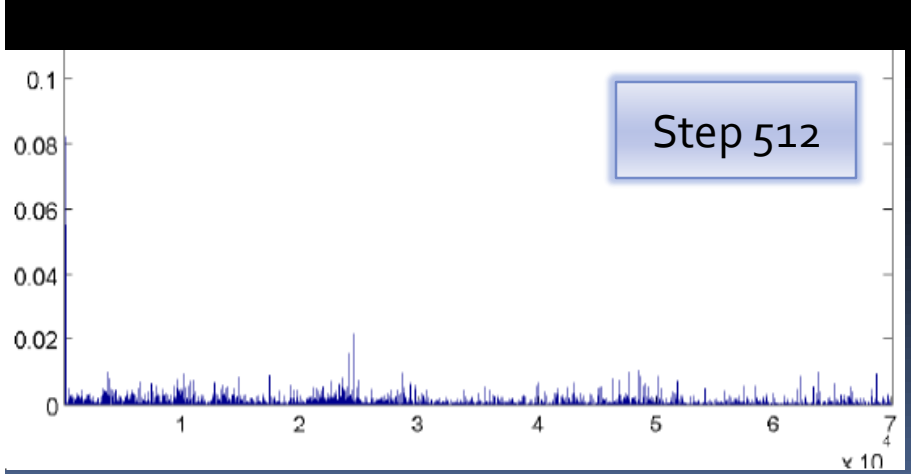
- It usually takes more than 1000 steps that we reach to the stationary distribution state
- Snapshots from Random Walks steps for a sample topic (Topic 83)
- Steps 1; 2^3; 2^5; 2^7; 2^9; and 2^{11}
- Scores of top ranked results:
- 0.3166, 0.1052, 0.0789, 0.0526 scores after 2048 steps.

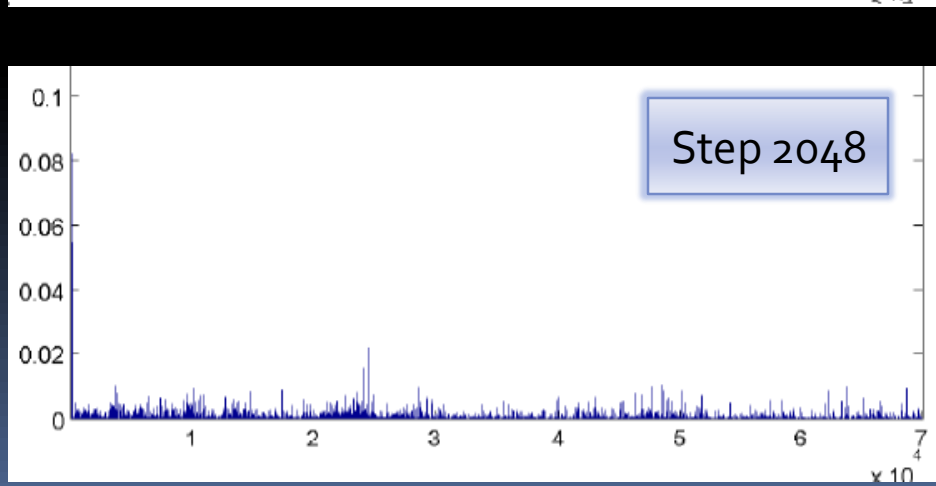

Step 1
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4


Step 8
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4


Step 32
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4


Step 128
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4


Step 512
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4


Step 2048
0.1
0.08
0.06
0.04
0.02
0
1
2
3
4
5
6
7
x 10^4

Precision analysis of RW and MH

Table 1: Performance result with Random Walks

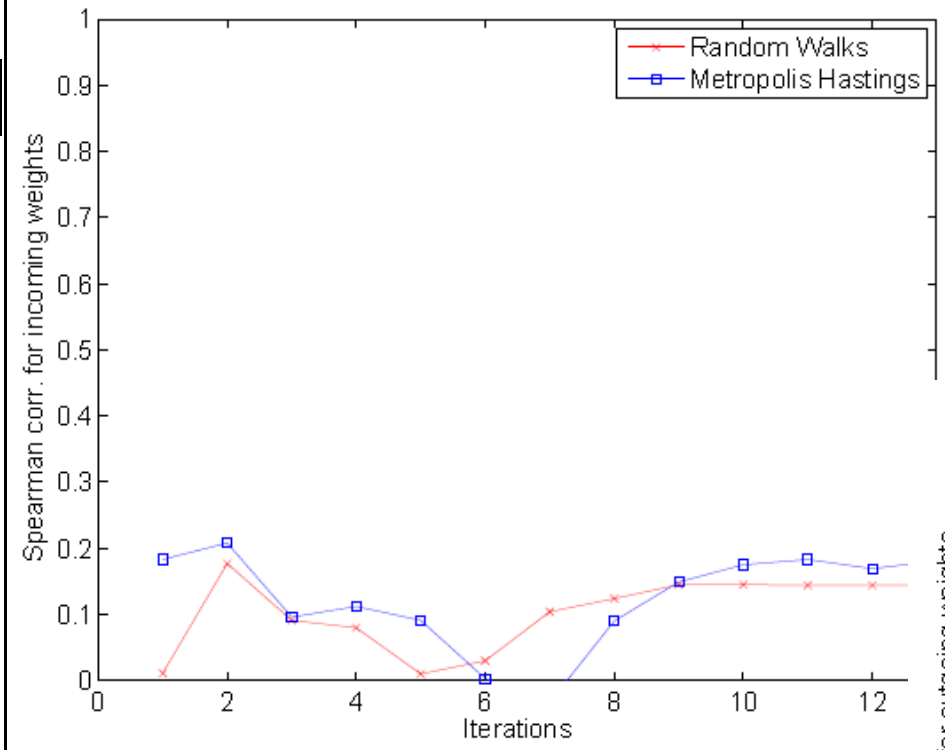
iter	*steps*	*p@10*	*r@10*	*p@20*	*r@20*
1	1	0.2267	0.0815	0.1767	0.1111
2	2	0.1571	0.0478	0.15	0.0866
3	4	0.1	0.0293	0.1067	0.0523
4	8	0.0857	0.0209	0.0643	0.0364
5	16	0.0133	0.0027	0.0333	0.0175
6	32	0.0	0.0	0.0	0.0
7	64	0.0	0.0	0.0	0.0
8	128	0.0	0.0	0.0	0.0
9	256	0.0	0.0	0.0	0.0
10	512	0.0	0.0	0.0	0.0
11	1024	0.0	0.0	0.0	0.0
12	2048	0.0	0.0	0.0	0.0
13	4096	0.0	0.0	0.0	0.0

Table 2: Performance Result with Metropolis-Hastings

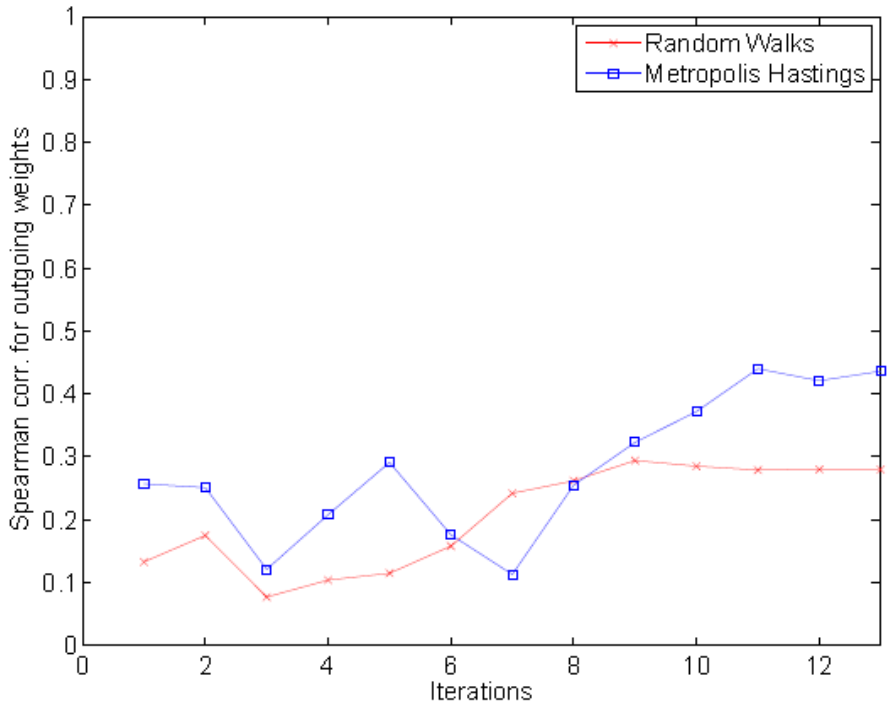
iter	*steps*	*p@10*	*r@10*	*p@20*	*r@20*
1	1	0.1958	0.0501	0.1479	0.0711
2	2	**0.216**	**0.065**	**0.16**	**0.0836**
3	4	0.175	0.0432	0.1562	0.0787
4	8	0.1333	0.029	0.1292	0.0582
5	16	0.1125	0.0284	0.0833	0.0354
5	32	0.068	0.0145	0.056	0.0189
7	64	0.032	0.0079	0.018	0.0081
8	128	**0.0042**	**0.0038**	**0.0021**	**0.0038**
9	256	0.0042	0.0038	0.0021	0.0038
10	512	0.0042	0.0038	0.0021	0.0038
11	1024	0.0042	0.0038	0.0021	0.0038
12	2048	0.0042	0.0038	0.0021	0.0038
13	4096	**0.0042**	**0.0038**	**0.0021**	**0.0038**

Correlation Analysis

- Between the top ranked results and the number of their neighbours
- What is the result bias to the number of nodes that a node is connected to.

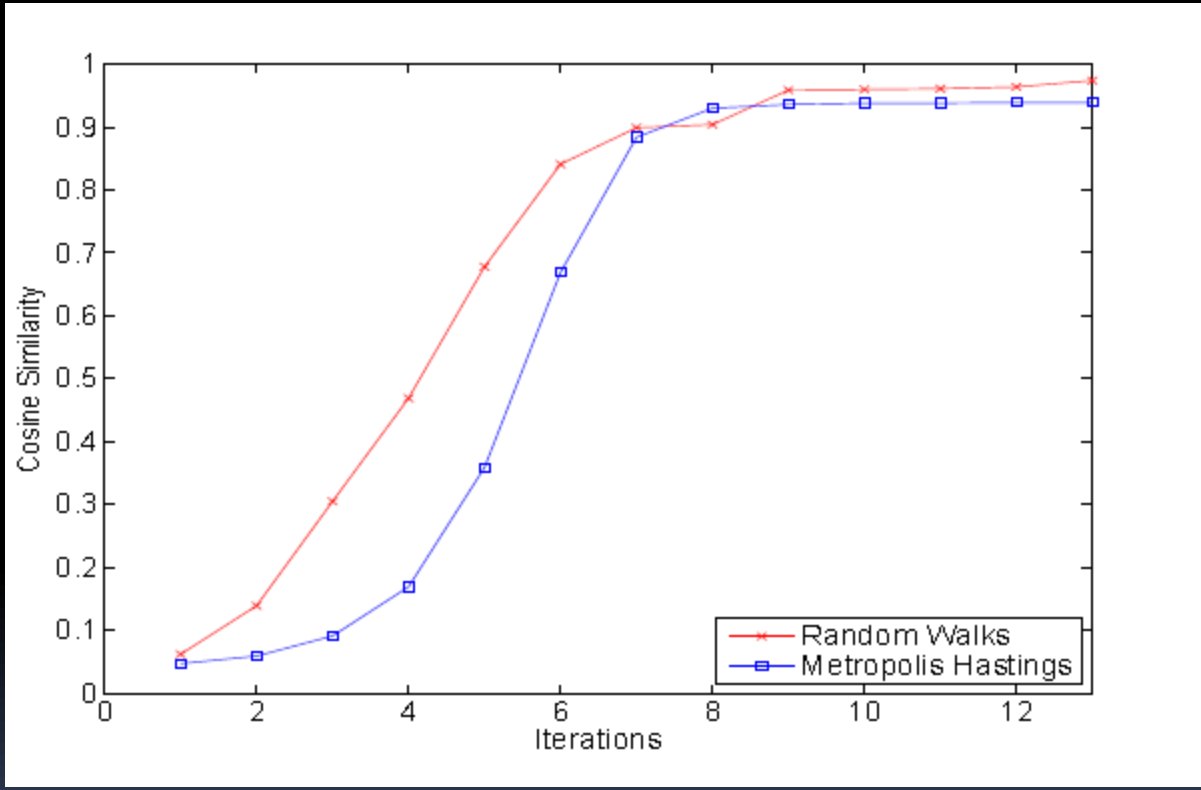


(a) Correlation between the top ranked resul
and the **incoming weights**



(b) Correlation between the top ranked results
and the **outgoing weights**

The rate initial score is overridden by the graph structure



Conclusion

- Compared the performance of using query-dependent and independent RW
- Higher precision results with query-dependent RW
- Can be improved by using better relevancy functions.
- Correlation analysis between the top ranked results and the No. of incoming/outgoing links of these nodes shows less dependency to the incoming link for MH.

What is a stationary distribution? Intuitively and Mathematically

What is a stationary distribution? Intuitively and Mathematically

- The stationary distribution at a node is related to the amount of time a random walker spends visiting that node.

What is a stationary distribution? Intuitively and Mathematically

- The stationary distribution at a node is related to the amount of time a random walker spends visiting that node.
- Remember that we can write the probability distribution at a node as
  - $x_{t+1} = x_t P$

What is a stationary distribution? Intuitively and Mathematically

- The stationary distribution at a node is related to the amount of time a random walker spends visiting that node.
- Remember that we can write the probability distribution at a node as
  - $x_{t+1} = x_t P$
- For the stationary distribution v_o we have
  - $v_o = v_o P$

Probability Distributions

- $x_t(i)$ = probability that the surfer is at node *i* at time *t*
- $x_{t+1}(i) = \sum_j$(Probability of being at node j)*Pr(j->i) $= \sum_j x_t(j) * P(j,i)$
- $x_{t+1} = x_t P = x_{t-1} * P * P = x_{t-2} * P * P * P = \ldots = x_0 P^t$
- What happens when the surfer keeps walking for a long time?

What is a stationary distribution? Intuitively and Mathematically

- The stationary distribution at a node is related to the amount of time a random walker spends visiting that node.
- Remember that we can write the probability distribution at a node as
  - $x_{t+1} = x_t P$
- For the stationary distribution v_o we have
  - $v_o = v_o P$
- Whoa! that's just the left eigenvector of the transition matrix!

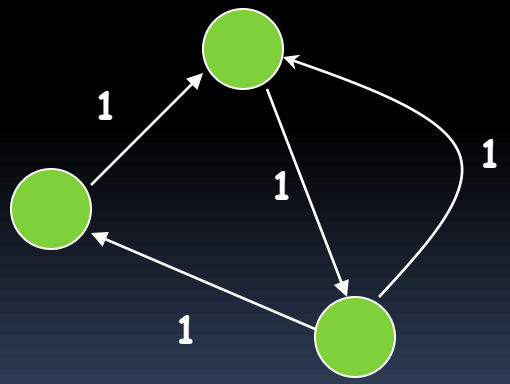
Definitions

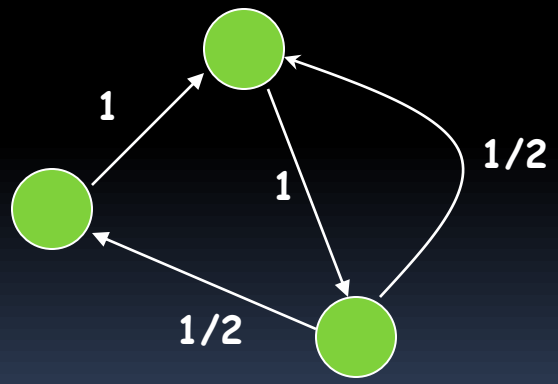
0	1	0
0	0	1
1	1	0

Adjacency matrix A

0	1	0
0	0	1
1/2	1/2	0

Transition matrix P



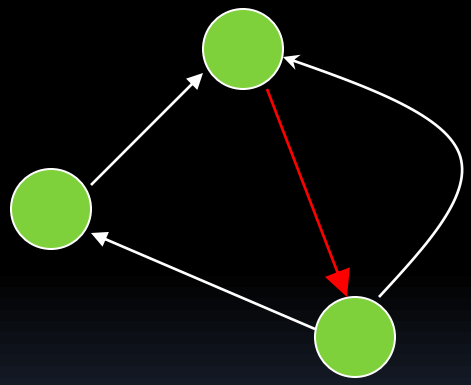


Interesting questions

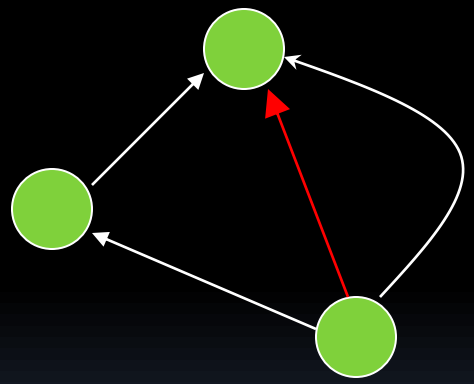
- Does a stationary distribution always exist? Is it unique?
  - Yes, if the graph is "well-behaved".
- What is "well-behaved"?
  - We shall talk about this soon.
- How fast will the random surfer approach this stationary distribution?
  - Mixing Time!

Well behaved graphs

- Irreducible: There is a path from every node to every other node.



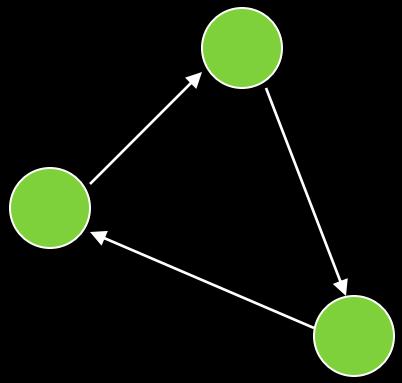
Irreducible



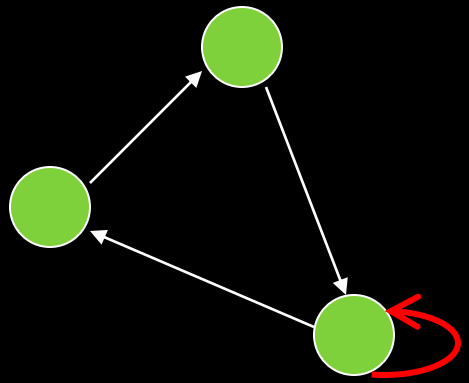
Not irreducible

Well behaved graphs

- Aperiodic: The GCD of all cycle lengths is *1*. The GCD is also called period.



Periodicity is 3



Aperiodic