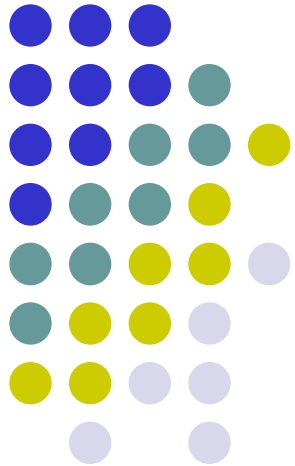
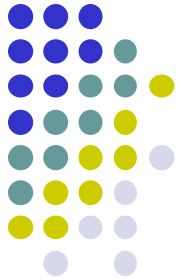
Web Algorithms – Web Search

Part 3: Topic-Specific PageRank

Eng. Fabio Persia, PhD

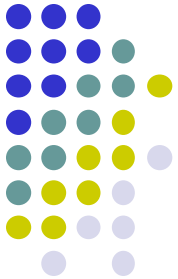


Topic-Specific PageRank



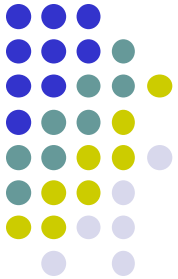
- **Instead of generic popularity, can we measure popularity within a topic?**
- **Goal:** Evaluate Web pages not just according to their popularity, but by how close they are to a particular topic, e.g. “sports” or “history”
- **Allows search queries to be answered based on interests of the user**
  - **Example:** Query “Trojan” wants different pages depending on whether you are interested in sports, history and computer security

Topic-Specific PageRank



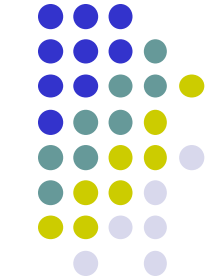
- Conceptually, we use a random surfer who teleports, with say 10% probability, using the following rule:
  - Selects a category (say, one of the 16 top level ODP categories) based on a query & user - specific distribution over the categories
  - Teleports to a page uniformly at random within the chosen category
- Sounds hard to implement: can't compute PageRank at query time!

Topic-Specific PageRank



- **Offline**: Compute pagerank for *individual* categories
  - Query independent as before
  - Each page has multiple pagerank scores – one for each ODP category, with teleportation only to that category
- **Online**: Distribution of weights over categories computed by query context classification
  - Generate a dynamic pagerank score for each page - weighted sum of category-specific pageranks

Influencing PageRank ("Personalization")

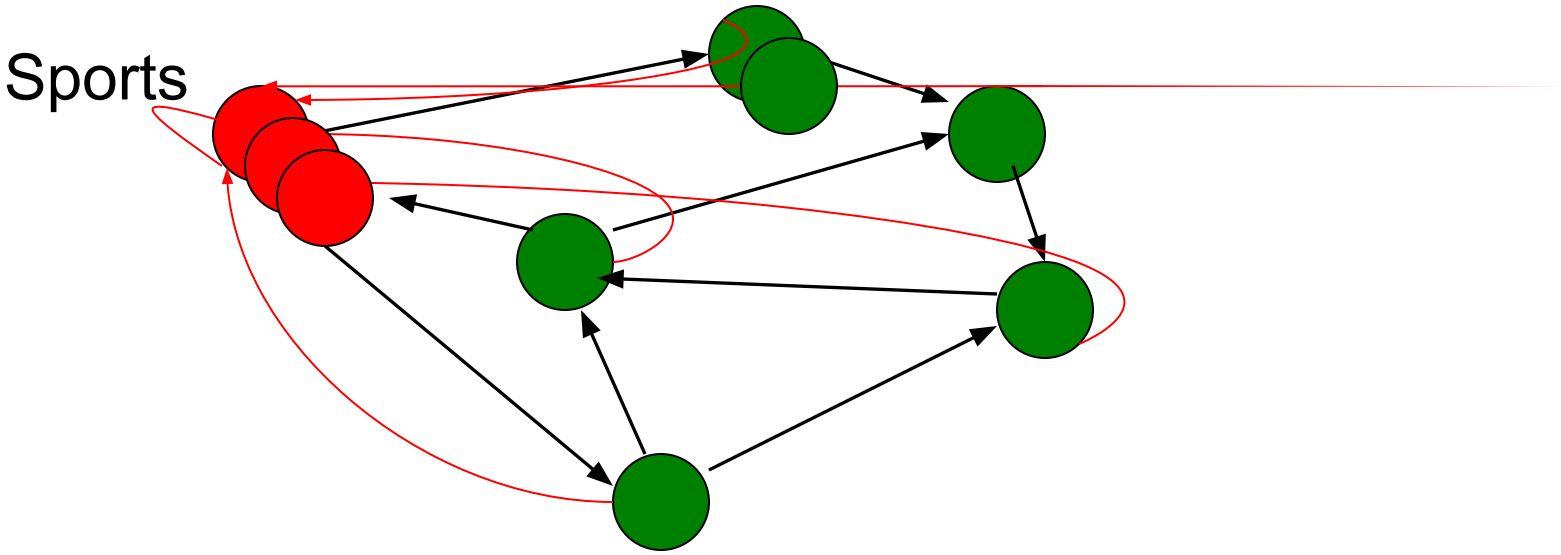


- Input:
  - Web graph *W*
  - influence vector **v**

    v : (page → degree of influence)

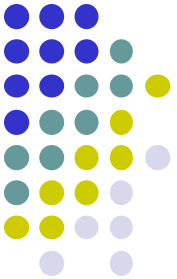
- Output:
  - Rank vector **r**: (page → page importance wrt **v**)
- **r** = PR(*W* , **v**)

Non-uniform Teleportation



Teleport with 10% probability to a Sports page

Interpretation of Composite Score

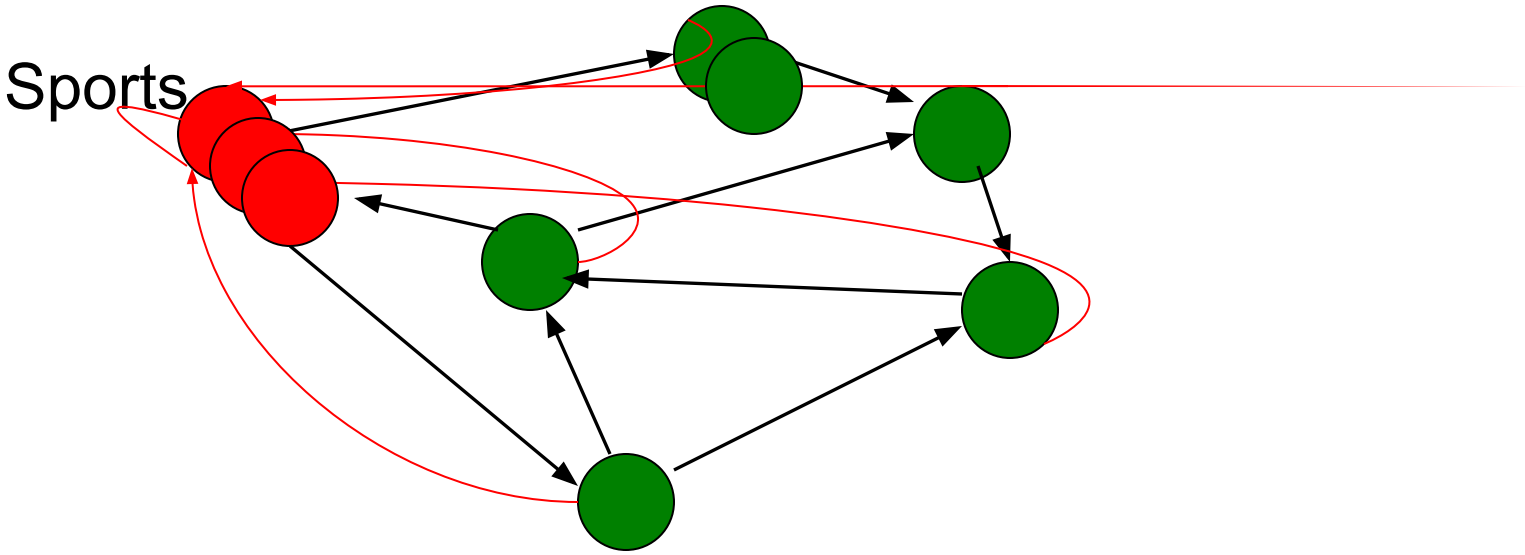


- For a set of personalization vectors $\{\mathbf{v}_j\}$

$$\sum_j [w_j \cdot PR(W, \mathbf{v}_j)] = PR(W, \sum_j [w_j \cdot \mathbf{v}_j])$$

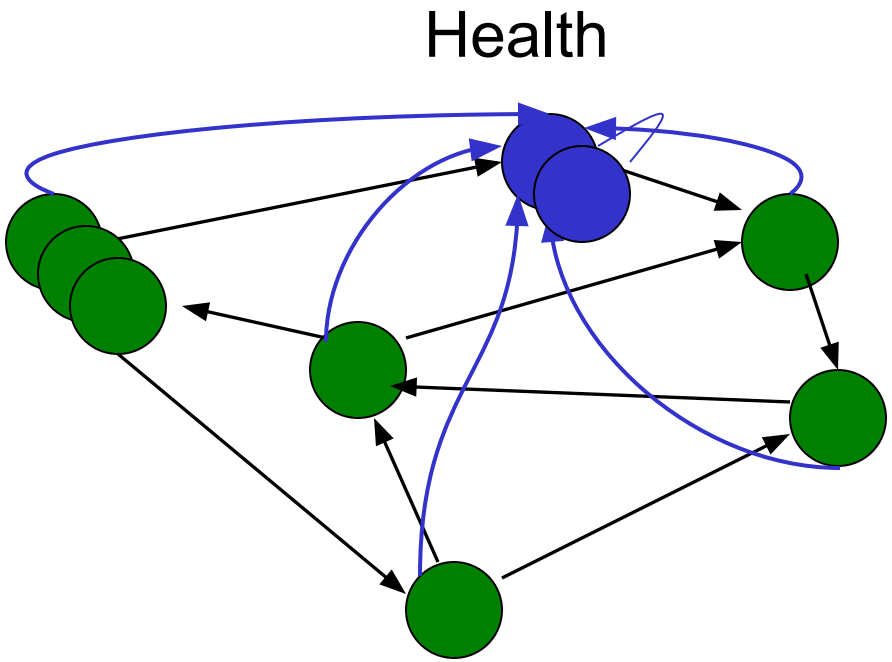
- Weighted sum of rank vectors itself forms a valid rank vector, because PR() is linear wrt $\mathbf{v}_j$

Interpretation



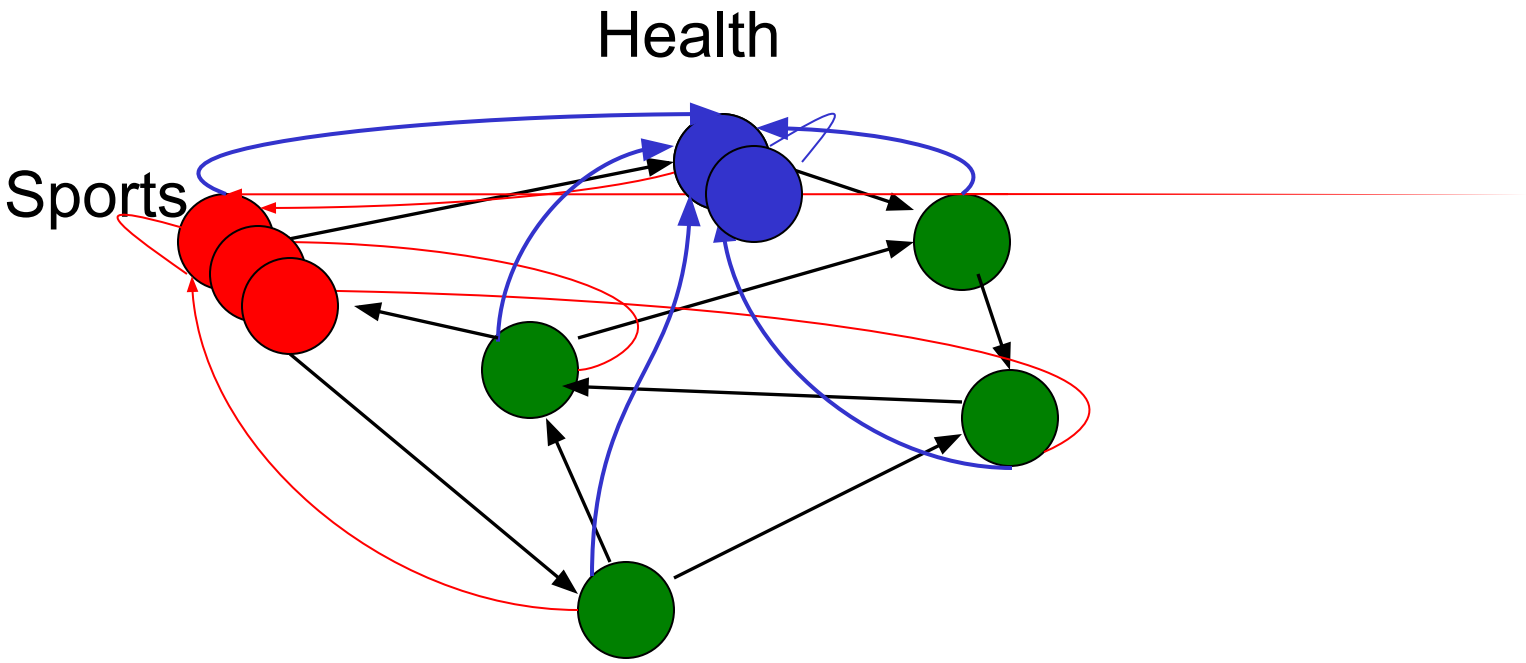
10% Sports teleportation

Interpretation



10% Health teleportation

Interpretation



pr = (0.09 PR_{sports} + 0.01 PR_{health}) gives you:
9% sports teleportation, 1% health teleportation