

---

---

# PageRank

Presenters:  
Ziting Shen & Rachel Xu

---

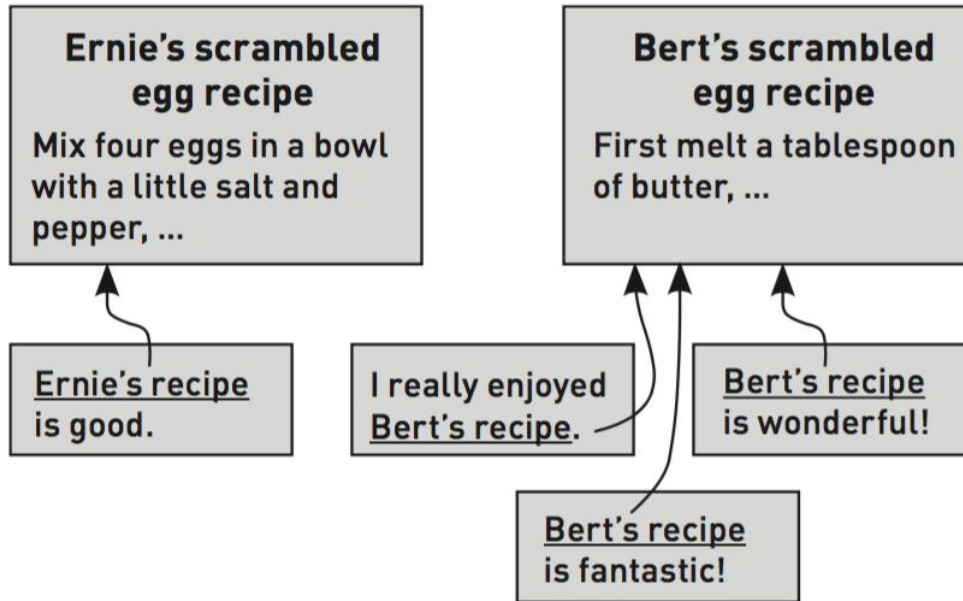
---

# Problem Description

- ❑ Algorithm chief inventor: Larry Page - co-founder of Google
- ❑ Input: web pages that match the query
- ❑ Output: rankings of these web pages, based on their desirability, credibility, relatedness, etc.
- ❑ Expectation:
  - ❑ as fast as possible
  - ❑ as “desirable” as possible



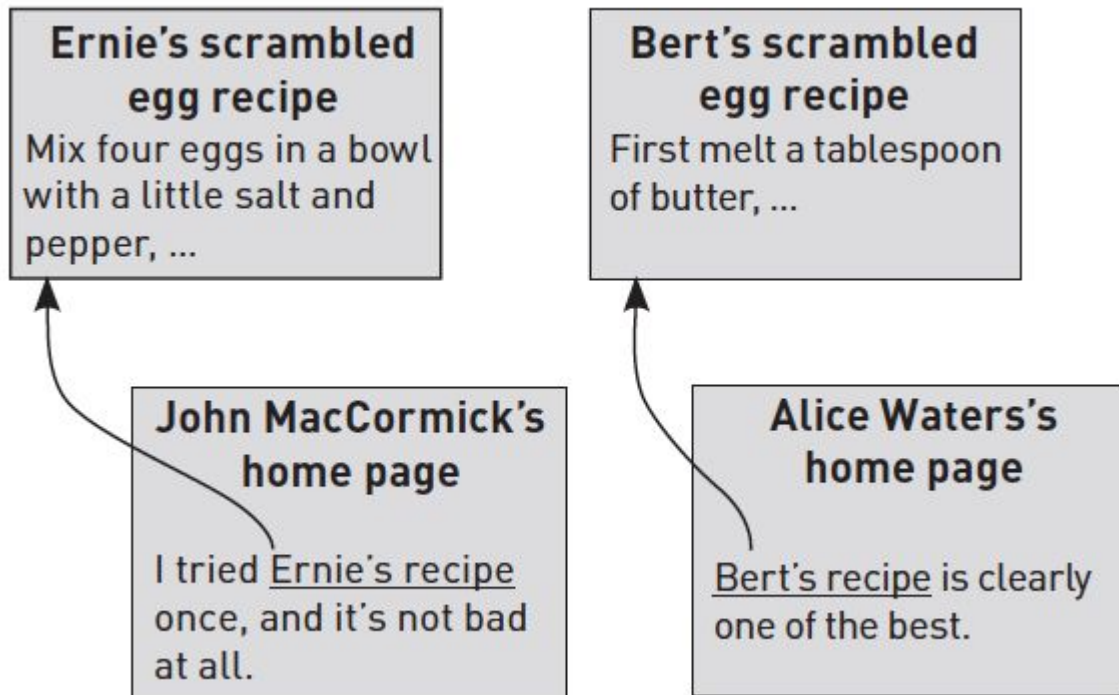
# The Hyperlink Trick



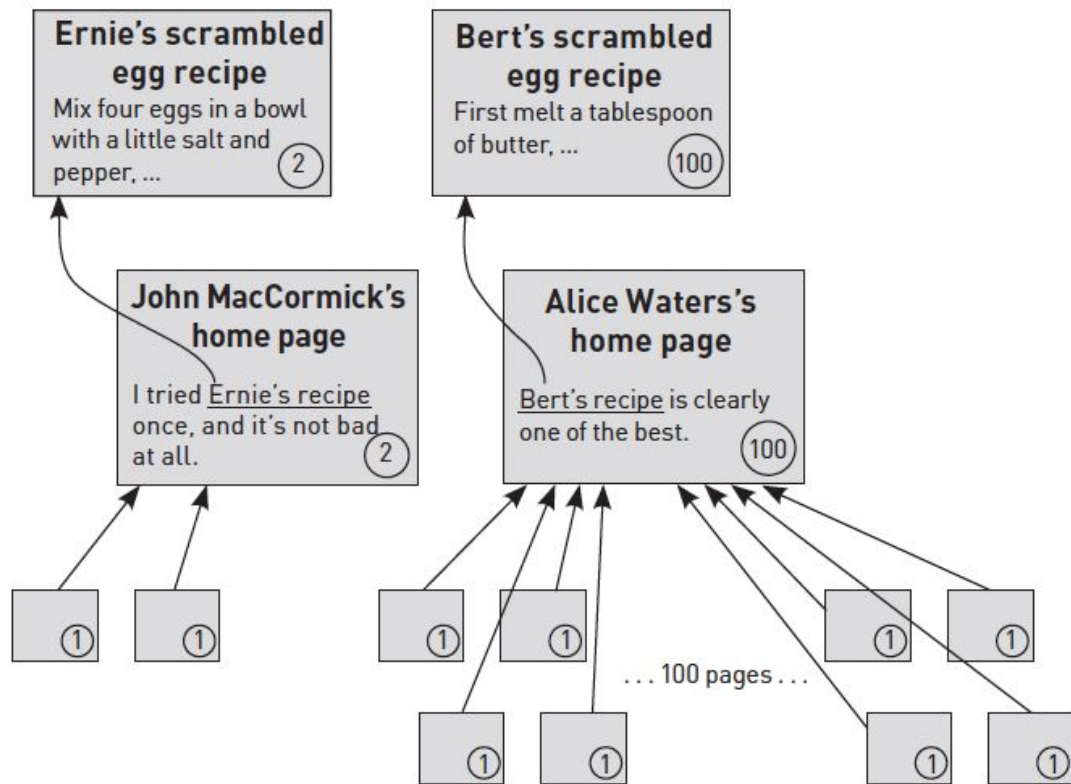
# The Hyperlink Trick

- ❑ How do we rank (evaluate the usefulness of) web pages?
  - ❑ More hits -> more helpful -> higher ranking
  - ❑ Hyperlinks -> visits
  - ❑ Directed graph: node -> web page, edge -> hyperlink
  - ❑ Computers don't understand the web pages, but can count hyperlinks
  - ❑ Number of incoming hyperlinks = indicator of usefulness
- ❑ Potential Problem: links to good or bad pages?

# The Authority Trick



# The Authority Trick

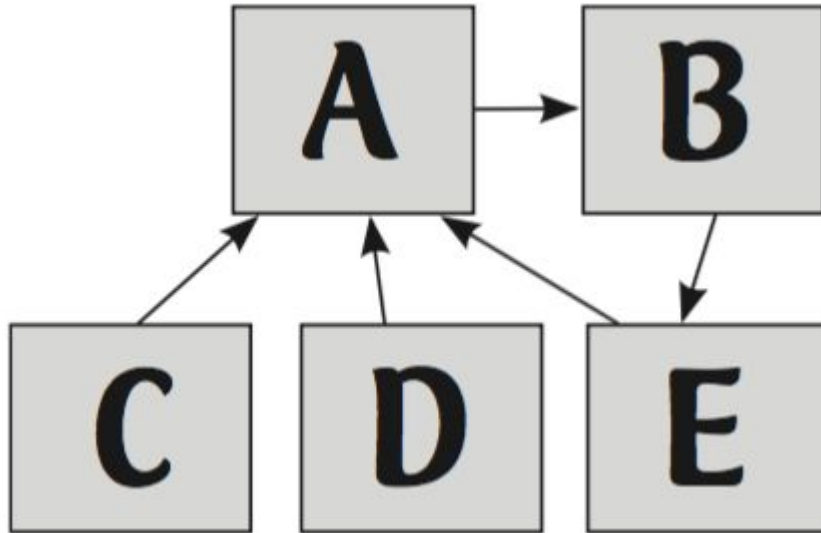


# The Authority Trick

- ❑ How to rank (evaluate the usefulness of) web pages?
  - ❑ Directed graph: node -> web page, edge -> hyperlink
  - ❑ Hyperlinks from the authorities -> more helpful -> higher ranking
  - ❑ Computers don't understand what's authoritative, but can count hyperlinks
  - ❑ Number of incoming hyperlinks = indicator of authoritativeness
  - ❑ More incoming hyperlinks -> more authoritative -> higher ranking for outgoing hyperlinks
- ❑ Potential Problem: more hyperlinks == authoritativeness?

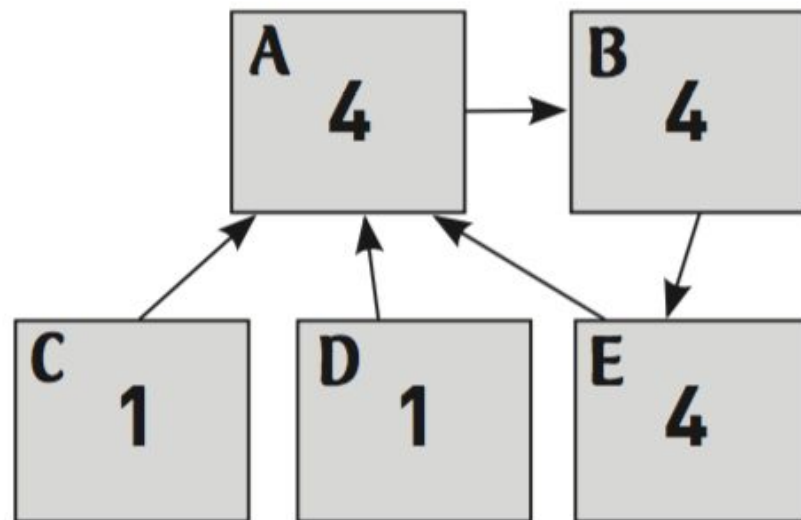
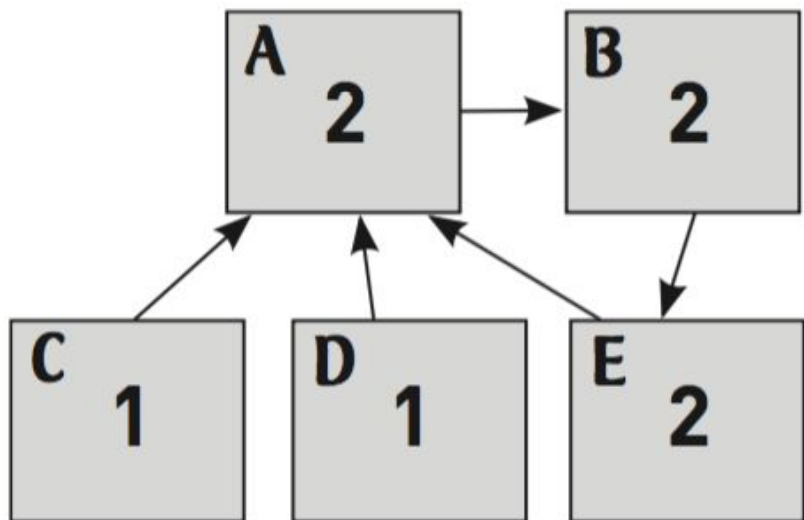
# The Random Surfer Trick

Major Problem with the Authority Trick: "cycles"



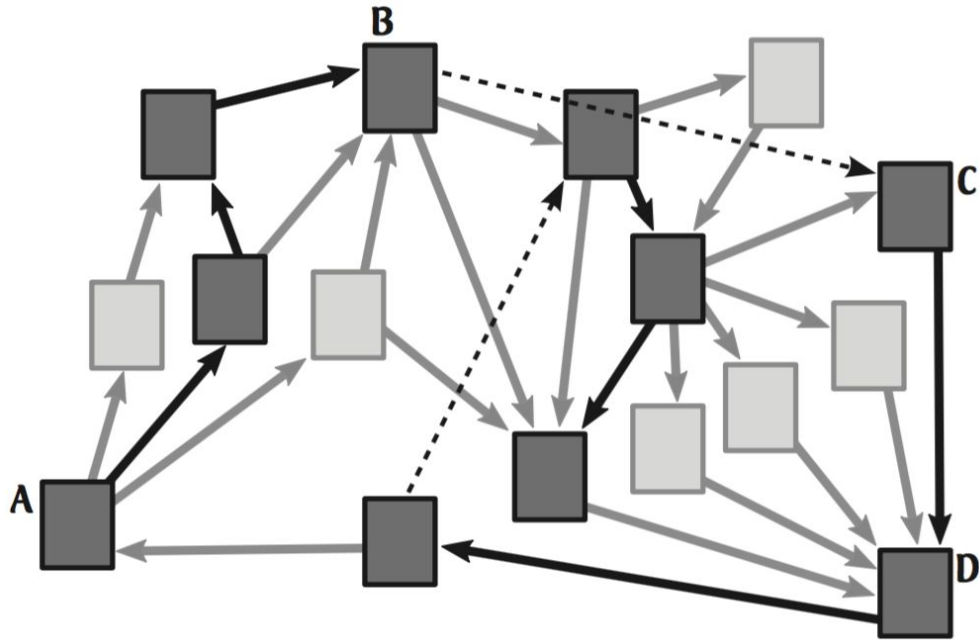


# The Random Surfer Trick



# The Random Surfer Trick

How do we solve chicken and egg problem?



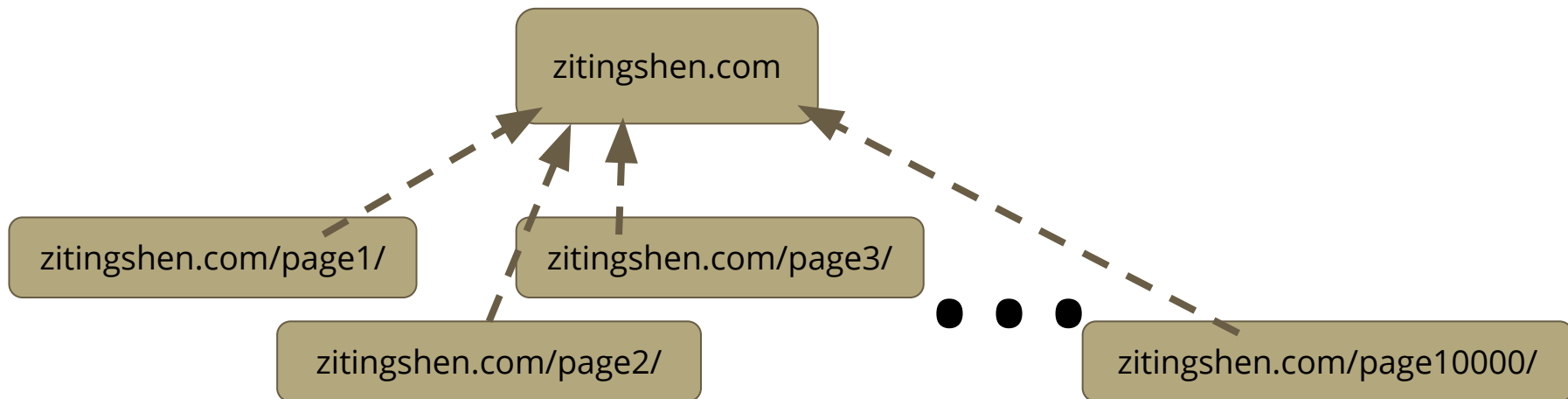
# The Random Surfer Trick

- ❑ Random surfer: Start at a random page -> select a random hyperlink -> go to a new page -> select a random hyperlink -> ... -> restart at a random page -> select a random page -> ... ..
- ❑ Restart Probability
- ❑ Surfer authority score = percentage of time the surfer spends visiting
- ❑ Larger number of random web pages -> improved accuracy
- ❑ Random Surfer incorporates Hyperlink and Authority



# Complicating Factor 1: Web Spam

- ❑ Generalization: link-based algorithms
- ❑ Assumption: hyperlinks confer legitimate authority
- ❑ Inadequacy: artificially inflation the ranking of certain web pages



# Complicating Factor 2: Efficiency

- ❑ Random simulation on the entire web is very slow and costly
- ❑ Mathematical techniques are used to reduce computational expense
- ❑ Multiple other algorithms contribute to the ranking
- ❑ The assumption holds true: hyperlinks confer authority

---

---

**Thank you =D**

---

---