proposal

title & problem addressing

- identify the most important page (the most likely page, which a random visitor would eventually arrive at) among UMJI official websites.

why is the problem important

- help freshmen in UMJI to get familiar with official websites
- the strategy can be applyed to other websites related to campus life (eg: canvas, i.sjtu.edu.cn)

why is linear algebra useful

- we introduce PageRank algorithm to rank the pages
- the probability of a random visitor staying at a certain page at time t can be represented by a stochastic row vector
- the main idea is applying linear orthogonal transform to the stochastic row vector

what area of linear algebra

- eigenvalue & eigenvector of Markov Chain
- fast computation of sparse linear system

sources

> Page, Larry, "PageRank: Bringing Order to the Web". Archived from the original on May 6, 2002. Retrieved 2016-09-11., Stanford Digital Library Project, talk. August 18, 1997 (archived 2002)
> Brin, S.; Page, L. (1998). "The anatomy of a large-scale hypertextual Web search engine". Computer Networks and ISDN Systems. 30 (1–7): 107–117. CiteSeerX 10.1.1.115.5930. doi:10.1016/S0169-7552(98)00110-X. ISSN 0169-7552. Archived from the original on 2015-09-27.
> Taher Haveliwala & Sepandar Kamvar (March 2003). "The Second Eigenvalue of the Google

> Matrix". Stanford University Technical Report: 7056. arXiv:math/0307056. Bibcode:2003math.....7056N. Archived from the original on 2008-12-17.
> Gianna M. Del Corso; Antonio Gullí; Francesco Romani (2005). "Fast PageRank Computation via a Sparse Linear System". Internet Mathematics. Lecture Notes in Computer Science. 2. pp. 118–130. CiteSeerX 10.1.1.58.9060. doi:10.1007/978-3-540-30216-2_10. ISBN 978-3-540-23427-2. Archived from the original on 2014-02-09.

graphs/tables

- table of raw data of outbound links
- table of the importance of websites (i.e. the probability of a random visitor staying on the website when $t \rightarrow \infty$)
- maybe we will include the screenshots of some of the important websites

meeting times per week

- 6 hours

how we divide the work

- Sun,Yi
  - Markov Chain
  - original PageRank algorithm
  - data collecting
- Zhu, Zhuoer
  - fast PageRank algorithm
  - statistics and probability
- Chen, Tianbo
  - coding of 2 types of PageRanking algorithm
  - designing the poster