Higher-Order Web Link Analysis Using Multilinear Algebra

1. What is the problem that the paper wants to solve? Why is it difficult (related works)?
   - Given a semantic graph of web pages, the paper wants to provide a list of authorities and hubs on certain topics (provided by the web pages anchor texts)
   - Other algorithms such as HITS suffer from "topic drift", that is the resulting authorities and hubs may not fit the original query
2. What is the solution? What is the main idea?
   - Kolda et al. introduce a new method called TOPHITS which, building on ideas from HITS, essentially models the semantic graph as a three-way tensor. Given this three-way tensor, one can compute the Parallel Factors (PARAFAC) decomposition to get an approximation of authority, hub and topic scores (similarly to SVD for HITS).
3. What is the result?
   - Results are very similar to HITS but for each website scores for semantic terms are included that help identify the page's topic and can make the search results more relevant to the query
4. What is the main novelty that enabled the solution?
   - The authors decided to include related topics in a third dimension and obtained a three-way tensor for which they were able to use the PARAFAC decomposition to compute authorities, hubs and topics
   - Additionally, the authors wrote a `sparse_tensor` class in MATLAB to process sparse tensors in order to compute the PARAFAC decomposition
5. What are the good aspects of the paper? Did you learn something from the paper?
   - Authority, hub and topic scores can be pre-computed offline, actual queries are fast even though PARAFAC decomposition for very large tensors may be slow
   - Power iteration runtime is comparable to HITS while TOPHITS gives us topical information as well
   - The authors leveraged semantic data (anchor texts) to achieve better results in search, furthermore allows for extraction of authorities on certain topics
   - Good and thoughtful explanations of any notation used in the paper included
6. What is the impact of the paper?
   - This paper is an important milestone in using tensor analysis and multilinear algebra for data analysis and to solve problems involving large amounts of data
7. Are there weaknesses/missing parts in the paper? How can you improve it?
   - Although they mention that the topic information may be used to address topic drift in HITS it is not clear how this algorithm compares to PageRank and whether search results are actually "better" (this is admittedly hard to quantify)
8. How can you extend the paper?
   - The computation of the PARAFAC decomposition for very large (millions of websites), very sparse tensors needs
   - As the authors mention: higher-order tensors with more data may provide additional benefits that can aid in finding the most relevant search results
9. How can you apply the technique to other data/problems?
   - One may add additional information to matrix information in a given problem and obtain a three-way (or higher-dimensional) tensor, information can be extracted using PARAFAC