ECE101: Exploring Digital Information Technologies **Fall 2022**
Week 7 Exercises **based on L12,L13; tested on Exam II**

Search and Recommendation Engines

These exercises are intended to help you master and remember the material discussed in lectures and explored in labs. In future semesters, we may make some or all of these exercises required, but for now they remain optional. We suggest that you do them as we go over the material, but you may also want to use them to review concepts before the exam.

We suggest that you use this version rather than the version without solutions to solve the problems before looking at the version with solutions. Many studies have shown that people often trick themselves into believing that they know how to solve a problem if they are presented with the answer before they try to solve the problem themselves.

1. [L12] Think back to the first part of our course and our discussion of the Domain Name Service (DNS). When your computer needs to find the IP address for whitehouse.gov, it makes a series of requests to the DNS hierarchy, eventually obtaining the desired IP address.

   Many research publications require the authors to select topics and/or keywords from a list published by a related professional organization. For example, an article might discuss "page ranking" or, more generally, "Internet search." One can then easily turn the relationship around: by filing each publication in a folder corresponding to each of the publication's keywords, one can allow someone who wants to learn about a topic to easily find all relevant publications.

   Now let's try to combine those ideas to replace search engines. Rather than crawling the web, each published web page (authors are not required to participate) could register keywords with a service like DNS—let's call it the Keyword Lookup Service, or KLS. KLS can organize the URLs by adding a given URL to a folder for each of the selected keywords. Then when someone wants to find relevant documents on the web, they can walk the KLS hierarchy to obtain the relevant list of URLs!

   Explain why such an approach is unlikely to work well in practice compared with the approach taken by Internet search engines.

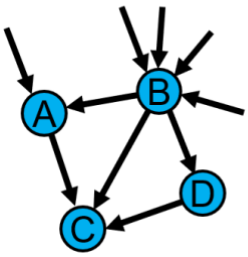
2. [L12] Explain the importance of indexing documents when operating a web search engine.

3. [L12] Several companies now offer integrated software platforms that allow your personal information to migrate freely from your desktop to your laptop to your phone and possibly even to a machine that you use at school or at a kiosk in an airport. Providing such conveniences requires that a company collect and retain that personal information on their servers, thus also enabling the company to tailor advertisements and web searches for you. For each of the following sources of information, give an example of how a company's use of a specific type of information collected from that context might benefit you.

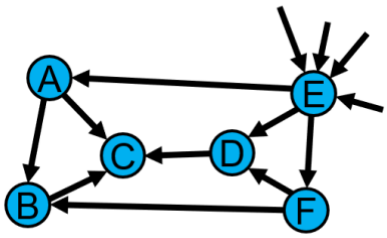
   A) web search
   B) online purchasing using a stored credit card
   C) email
   D) messaging
   E) social media
   F) phone calls
   G) videoconferencing

4. [L12] Each of the following graphs comes from a web search—all incoming arcs are shown. For each graph, rank the pages (nodes) in decreasing order of reputation according to page rank. (We don't expect you to do the full computation—just compare the number of incoming arcs. To break ties, use a second level of incoming arcs (how many nodes point to nodes that point to a page).

Graph #1



Graph #2



5. [L13] Recommendation systems work best for things that any given person chooses frequently. For example, one person watches many movies, eats food every day, and engages regularly in their hobbies. In contrast, one person may only buy a washing machine every ten or twenty years. Collaborative filtering is thus practically useless for such types of purchases.

   Content-based filtering, however, may play a role. To use the same example, washing machines have a rich feature space, including size, noise, cost, power, style, color, capabilities, and so forth. But a given user typically has no purchase history in the last decade. How can the feature space of washing machines be used to provide guidance on purchasing in a more interactive way? *(Hint: consider first applying clustering from our machine learning discussion in Week 8 to an assortment of products within the feature space.)*

6. [L13] For each of the following categories, suggest three possible dimensions for a feature space in which objects from the category could be placed in order to support recommendations.

   A) university Bachelors degree programs
   B) picture frames
   C) carbonated beverages