Distributed Information Systems: Spring Semester 2016

Quiz 5

Student Name: _____ Date: 12 May 2016

Student ID: _____ Time: 11:15AM to 11:30AM

Total number of questions: 8

Each question has a single answer!

1. Which of the following is **true** in the context of Girvan-Newman method for community detection?

   ☐ *a*) A depth first search algorithm is used in order to compute edge betweenness.

   ☐ *b*) Running time of the algorithm is $O(e)$, where e is the number of edges.

   ☒ *c*) Betweenness scores can take non-integer values.

   ☐ *d*) The method proceeds by removing edges with the lowest betweenness to form the communities.

2. The people of Washington decided to split their city into smaller, connected cities. They decided to use Girvan-Newman's community detection method in order to determine how to perform the split. They took houses as the nodes and streets connecting the houses as the edges. Which of the following is **true**:

   ☐ *b*) Some houses might end up as registered in two different cities.

   ☒ *b*) Streets with the highest betweenness will be the ones connecting the new cities.

   ☐ *c*) Houses on the same street will always end up in the same city after the split happens.

   ☐ *d*) None of the above

3. Which of following is **false** in the context of recommender systems?

   ☐ *a*) Matrix Factorization is a latent model.

   ☒ *b*) Recommender systems based on collaborative filtering measure similarity based on common attributes of items or users.

   ☐ *c*) Both collaborative filtering and content-based recommender systems suffer from the cold-start problem.

   ☐ *d*) Content-based recommendation is in general more scalable than collaborative filtering.

4. Which of following is **false** about the graph models?

   ☐ *a*) In random graphs, nodes are connected with uniformly random probability.

   ☐ *b*) Regular graphs have both a high clustering coefficient and a high diameter.

   ☐ *c*) In small-world networks built by Watts and Strogatz from regular networks, nodes have a fixed number of links in addition to random links.

   ☒ *d*) Random graphs are a realistic model for real-world social networks.

5. Which of the following is **true** in the context of community detection?

- ☐ *a*) The Girvan-Newman method uses an agglomerative algorithm for forming communities.
- ☐ *b*) Edge betweenness measures the total length of the shortest paths passing over an edge.
- ☒ *c*) The Louvain method is significantly more scalable than the Girvan-Newman method.
- ☐ *d*) Each weak community is also a strong community.

6. Which of the following is a **disadvantage** of content-based recommendation?

- ☐ *a*) It is only applicable in the presence of a community of users.
- ☒ *b*) Extracting features can be challenging for some types of media items.
- ☐ *c*) Feature extraction for all items has to be performed again whenever a new item is added to the collection.
- ☐ *d*) It tends to suggest a lot of unexpected recommendations, rather than more similar items.

7. An online dating platform provides recommendations about users you might be interested in. Which algorithm can make these recommendations?

- ☐ *a*) A user-based collaborative filtering that uses ratings from users with preferences similar to yours.
- ☐ *b*) An item-based collaborative filtering based on the users you already liked
- ☐ *c*) A content-based recommender systems using attributes of the users (what they like/dislike, etc)
- ☒ *d*) All of the above algorithms can make these recommendations.

8. For a graph with n nodes, which of the following is **correct**:

- ☐ *a*) The shortest path in a Kleinberg small world graph has always expected length $O(\log n)$.
- ☐ *b*) The shortest path in a clique, a fully connected graph, has always expected length $O(\log n)$.
- ☒ *c*) The shortest path in a random graph has always expected length $O(\log n)$.
- ☐ *d*) None of the above statements is correct.