

Social Network Analysis

1. What is Social Network Model and how is it represented? (10 pts)

A social network model is a set of actors and the interactions between them. It is successfully represented as a graph. Actors are nodes and interactions or relationships are links or edges. Actors might be connected by different kinds of ties and relations (eg: Friendship, Hate, Trust, etc.)

2. Define: (15 pts)

Degree, Closeness/Betweenness centrality, Cliques, Size, Out-degree, Eigen-vector.

- **Degree:** Sum of connections (sum of the weights of connections in case of weighted graphs) from or to an actor.
- **Closeness centrality:** Distance of one actor to all others in the network.
- **Betweenness centrality:** The number of shortest paths that passes through an actor.
- **Eigen-vector:** measures the importance of an actor.
- **Cliques:** Sub-set of actors. More closely tied to each other than to actors who are not part of the sub-set. Often, you first find the clique (or a densely connected subgraph) and then try to interpret what the clique is about.
- **Size:** Number of nodes.
- **Out-degree:** Sum of connections from an actor to others.

Other definitions covering other definitions that may be asked:

- **Social roles:** Defined by regularities in the patterns of relations among actors.
- **Density:** Number of ties that are present or the amount of ties that could be present.
- **In-degree:** Sum of connections to an actor.
- **Walk:** A sequence of actors and relations that begins and ends with actors.
- **Geodesic distance:** The number of relations in the shortest possible walk from one actor to another.
- **Maximum flow:** The amount of different actors in the neighborhood of a source that lead to pathways to a target.

- **Social networks:** links of actors and their relationships usually represented as a graph or network
- **Social networking:** the actual implementation of social networks in the digital world or media. A social network service focuses on building and reflecting social networks or social relations among people, e.g., who share interests and/or activities. A social network service essentially consists of a representation of each user (often a profile), his/her social links, and a variety of additional services. Most social network services are web based and provide means for users to interact over the internet, such as e-mail and instant messaging.
- **SEO (Search Engine Optimization):** involves altering website code, content, and presence in order to increase rankings in search engines
- **Crawling:** The purpose of the crawl is to identify relevant pages for indexing and assess whether they have changed. Crawling is performed by robots (bots) which are also known as spiders. These access web pages and retrieve a reference URL of the page for later analysis.
- **Indexing:** An index is created to enable the search engine to rapidly find the most relevant pages containing the query typed by the searcher. Rather than searching each page for a query phrase, a search engine ‘inverts’ the index to produce a lookup table of the documents containing particular words.
- **Ranking & Scoring:** The indexing process has produced a lookup of all the pages that contain particular words in a query, but they are not sorted in terms of relevance. Ranking of the document to assess the most relevant set of documents to return in the Search engine result pages (SERPs) occurs in real time for the search query entered. First, relevant documents will be retrieved from a runtime version of the index at a particular data center, then a rank in the SERPs for each.
- **Query request and Results Serving:** The familiar search engine interface accepts the searcher’s query. The user’s location is assessed through their IP address and the query is then passed to a relevant data center for processing. Ranking then occurs in real time for a particular query to return a sorted list of relevant documents, and these are then displayed on the Search Results Page.

- **While Label:** Offer some group-building functionality, which allows users to form mini-communities within sites (eg: Ning) but seems to be losing popularity.
- **Multi-user Virtual Environment:** allow users to interact with each other's avatars. Users have profile cards; their functional profiles are the characters they customize or build and control. Friends' lists are usually private and not publicly shared or displayed (eg: World of Warcraft).
- **Mobile Social Networking:** Many social networking sites (eg: X – Formerly Twitter) offer mobile phone versions of their services, allowing members to interact with their friends via their phones. mobile-led and mobile-only communities include profiles and media-sharing just as with web-based social networking services.
- **SNA (Social Network Analysis):** a method used to analyze connections between individuals, groups, or institutions, helping us understand how network structures influence behavior and interactions across various fields like politics, sociology, or transportation.
- **PageRank (PR):** is an algorithm used by Google Search to rank web pages in their search engine results. PageRank is a way of measuring the importance of website pages. PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.

Additional notes (read only):

- A social network service focuses on building online communities of people who share interests and/or activities, or who are interested in exploring the interests and activities of others. Most social network services are web based of ways for users to interact, such as and provide a variety e-mail and instant messaging services.
- Social Search engines are an important web development which utilize the popularity of social networking services. There are various kinds of social search engine, but sites like Wink and Spokeo generate results by searching across the public profiles of multiple social networking sites, allowing the creation of web-based dossiers on individuals. This type of people search cuts

across the traditional boundaries of social networking site membership, although any data retrieved should already be in the public domain.

3. Profile Based, Content Based and Micro Blogging are three types of social networking, what are the specifications of each one.

(10pts)

- **Profile Based:** Primarily organized around members' profile pages – pages that mainly consist of information about an individual member, including the person's picture and details of interests, likes and dislikes (eg: Facebook, Instagram).
- **Content Based:** In these services, the user's profile remains an important way of organizing connections, but plays a secondary role to the posting of content (eg: YouTube for video sharing).
- **Micro Blogging:** Allows you to publish short (140 characters, including spaces) messages publicly or within contact groups. These services are designed to work as mobile services, but are popularly used on the web as well (eg: X (formerly Twitter)).

4. Define HITS.

(5pts)

HITS (**Hyperlink Induced Topic Search**) is a method used in website analytics and search algorithms where a **hit** refers to a request made to a web server to download a specific file or resource; unlike a **visit**, which counts a user's access to a webpage, multiple hits can occur during one visit because each individual file (images, scripts, etc.) counts as a separate hit, often making hits much higher in number than visits. HITS is also applied in areas like the Clever model, fine-grained ranking, and query-sensitive retrieval.

Additional notes (read only) :

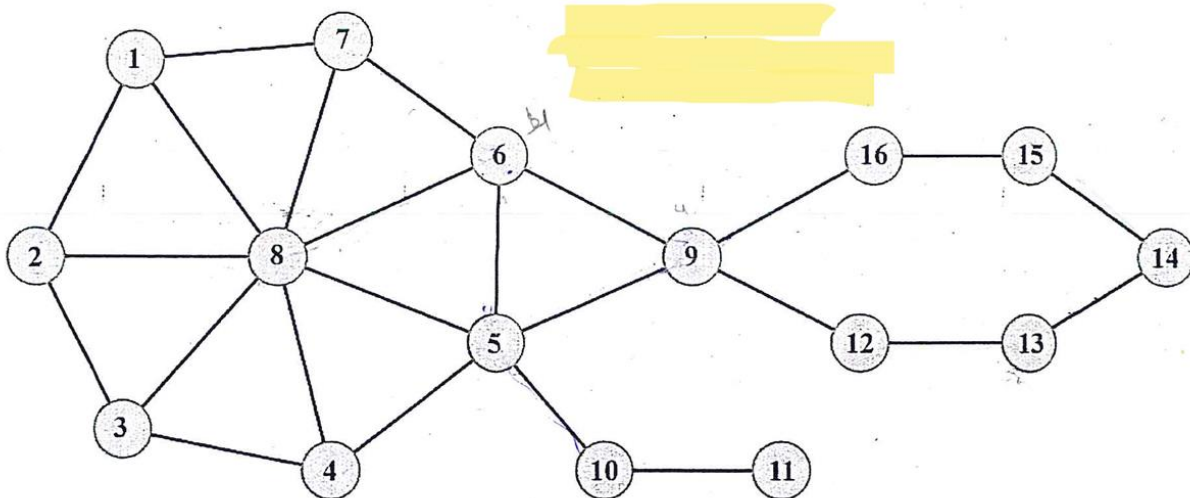
- Flow of Models: HITS & Google Each page is a node without any textual properties. Each hyperlink is an edge connecting two nodes with possibly only a positive edge weight property. Some preprocessing procedure outside the scope of HITS chooses what sub-graph of the Web to analyze in response to a query.
- High prestige $\Leftrightarrow$ good authority
High reflected prestige $\Leftrightarrow$ good hub

- Relation between HITS, PageRank and LSI - HITS algorithm = running SVD on the hyperlink relation (source, target) - LSI algorithm = running SVD on the relation (term, document). - PageRank on root set R gives same ranking as the ranking of hubs as given by HITS

5. In the following Picture:

(15pts)

- 1- Which Node has the highest Degree.
- 2- Which Node has the highest Betweenness.
- 3- Which Node has the highest Closeness.



1. Degree Centrality

Node 8 : It has the highest number of direct connections (7 edges), making it the most locally connected node in the graph.

2. Betweenness Centrality

Node 9 : It acts as the sole bridge between the dense left cluster and the right-side chain, appearing on most shortest paths across the graph.

3. Closeness Centrality

Node 5 : It sits at the center of the graph and can reach all other nodes in the fewest average number of steps.

6. Consider the following dataset and we will find frequent itemsets.

Note that minimum support count is 2.

(25pts)

TID	Items
T1	I1, I2, I5
T2	I2, I4
T3	I2, I3
T4	I1, I2, I4
T5	I1, I3
T6	I2, I3
T7	I1, I3
T8	I1, I2, I3, I5
T9	I1, I2, I3

Answer:

Min Sup = 2 / 9			Frequent item sets:		
h=1	Item	Support			
	I1	6	✓	{I1}, {I2}, {I3}, {I4}, {I5},	
	I2	7	✓	{I1, I2}, {I1, I3}, {I1, I5}, {I2, I3},	
	I3	6	✓	{I2, I4}, {I2, I5}, {I1, I2, I3}, {I1, I2, I5}	
	I4	2	✓		
	I5	2	✓		
h=2	I1, I2	4	✓		
	I1, I3	3	✓		
	I1, I4	1	✗		
	I1, I5	2	✓		
	I2, I3	4	✓		
	I2, I4	2	✓		
	I2, I5	2	✓		
	I3, I4	0	✗		
	I3, I5	1	✗		
	I4, I5	0	✗		
h=3	I1, I2, I3	2	✓		
	I1, I2, I5	2	✓		

Additional exercise, same question and information as number 6:

Transactions	Items
T1	a, b, c
T2	a, b
T3	a, b, d
T4	b, e
T5	b, c, e
T6	a, d, e
T7	a, c
T8	a, b, d
T9	c, e
T10	a, b, d, e
T11	a, b, c, e

Answer:

$\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{a, b\}, \{a, c\}, \{a, d\}$
 $\{a, e\}, \{b, e\}, \{b, d\}, \{b, c\}, \{c, e\}, \{d, e\}$
 $\{a, b, c\}, \{a, b, d\}, \{a, b, e\}$
 $\{a, d, e\}, \{b, c, e\}$