# Social networks

-The Ego networks are essentially networks centered around one particular node. Such networks are formed by taking one node, and finding all the vertices that are directly connected to it while also finding the connections between those vertices. We have used two ego networks, namely the “Ego-Facebook” dataset which contains 4039 nodes and 88233 edges as it combines 10 ego networks, and “Ego-Twitter” dataset which contains 81306 nodes and 2420765 edges. The Facebook ego networks were collected by Facebook users willingly using an app and twitter information was collected using web crawlers.

- Another type of network is formed between pages instead of account-holders. The “Gemsec-Facebook” datasets consist of networks of 8 different categories- Artists (50515 nodes and 819306 edges), Athletes (13866 nodes and 86858 edges), Company (14113 nodes and 52310 edges), Government (7057 nodes and 89455 edges), New sites (27917 nodes and 206259 edges), Politicians (5908 nodes and 41729 edges), Public figures (11565 nodes and 67114 edges), and Tv shows (3892 nodes and 17262 edges). Each individual dataset consist of verified Facebook pages of that category, and edges exist between pages that like each other.

-Twitch is a website on which people can livestream, and it is used mostly by gamers. The “Musae-twitch” datasets consists of 6 networks – DE (9498 nodes and 153138 edges), ENGB (7126 nodes and 35324 edges), ES (4648 nodes and 59382 edges), FR (6549 nodes and 112666 edges), PTBR (1912 nodes and 31299 edges), RU (4385 nodes and 37304 edges) - divided by the languages used by the streamers. The edges represent friendships between the streamers. The data was collected on May 2018.

-Last.fm is a website that has been providing music services since 2003. The “Feather-lastfm-social” dataset contains a social network of LastFM users which was collected in March 2020. The nodes represent users residing in Asian countries and edges represent mutual follow relationships. It has 7624 nodes and 27806 edges.

# Networks with ground-truth communities

-A network with ground-truth communities available aid in research as they provide a basis for comparison when communities are detected or identified using any algorithm. Such a networks can be constructed from the data available at the DBLP computer science bibliography website. The “Com-DBLP” dataset has 317080 edges and 1049865 nodes and it represents a network of authors. The nodes represent authors and an edge exists between two nodes if they have published at least one paper together.

# Communication Networks

-Communication networks can be constructed by collecting email data from institutes. Such networks are the “Email-EuAll” and “Email-enron” networks. “Email-EuAll” has 265214 nodes and 420045 edges, and was collected from an European institute from October 2003 to May 2005. The nodes represent email address and a directed edge signifies that source node sent at least one email to target node. “Email-enron” has 36692 nodes and 367662 edges and was collected, and made public, from the Enron Corporation when it was being investigated. This network is undirected and contains an edge between nodes if any email was exchanged between them.

# Collaboration networks

-The Arxiv collaboration networks were collected from the e-print arXiv websites. The edges represent collaborations between authors, so if one author collaborated with another an undirected edge exists between them. The data consists of papers published between January 1993 to April 2003. The “Ca-AstroPH” dataset has 18772 nodes and 396160 edges, and represents collaborations in papers of the Astro Physics category. The “Ca-HepPH” dataset has 12008 nodes and 237010 edges, and represents collaborations in papers of the High Energy Physics - Phenomenology category.

-The Arxiv citation networks were also collected from the e-print arXiv database. These datasets represent which papers cite each other within this database, and if one paper cites another, a directed edge is drawn from the former to the latter. Information regarding cited papers that do no exist within the database is not present in the networks. The data is collected from papers published between January 1993 and April 2003. The “cit-HepPH” dataset contains 34546 nodes and 421577 edges and consists of papers submitted to the High Energy Physics- Phenomenology category. The “cit-HepTH” dataset contains 27770 nodes and 352807 edges and consists of papers submitted to the High Energy Physics- Theory category.

# Co purchasing networks

-The Amazon product co-purchasing networks were collected by crawling the website and collecting information from the “Customers Who Bought This Item Also Bought” feature. A directed edge exists from one product to another if the former is frequently bought companying the latter. The “Amazon0302” dataset was collected on 2nd March 2003 and has 262111 nodes and 1234876 edges. The “Amazon0601” dataset was collected on 1st June 2003 and has 403394 nodes and 3387388 edges.

# Internet peer-to-peer networks

-The hosts in a peep-to-peer sharing system also form networks. Such networks were constructed from the Gnutella file sharing network by taking 9 snapshots across a few days in August 2002. The ones used in this research are : “P2p-Gnutella04” (10876 nodes and 39994 edges) which was collected on 4th August, “p2p-Gnutella30”( 36682 nodes and 88328 edges) which was collected on 30th August, “p2p-Gnutella31” (62586 nodes and 147892 edges) which was taken on 31st August.

# Signed networks

Unlike other social networks where edges between nodes only signify the existence of a relation, but not the nature, signed networks are able to convey both this information. Such a network is the “Soc-sign-Slashdot090221” which has 82140 nodes and 549201 edges and shows whether two connected nodes are friends or foes. However, in this research the sign is not taken into account. The network was collected in February 2009.