



# Project Supplement Material

Muhammad Qasim Pasta

## Network Extraction

One of the major tasks in network analysis is gathering data of the domain we are interested in. There are a number of techniques which can be used for this. In some cases, there is no prior data available like if we want to map the friendship network of students of a campus (considering there is no social network site exists for campus). In such cases we have to perform a survey in order to gather desired data.

In other cases, we may have some kind of data repository in form of reports, catalogues, or directories etc. When there is a repository available, the next task is to generate a network by creating an edge list or adjacency matrix. There are a number of such repositories also available in digital form on the internet like online social networks etc. One can write a script or crawler to extract information from these digital libraries.

There is no doubt that extracting a network from a repository is a tedious and a side job. But it is essential and critical for network analysis. We can summarize the data gathering task as follows:

- Domain Identification
- Identify whether there is any data repository available for the domain, if no then go for survey.
- If there is a repository available then we have to decide which information will be represented by a node and what kind of relationship can be represented by an edge.
- If there is a physical repository available then convert that repository into a digital format like spreadsheet or directly into an edge list.
- If there is an online repository, then devise a strategy to extract information by writing a script or a crawler.

It is important to note that before proceeding to generate a network a person must identify the key questions which can be answered by generating that network. If the questions are such simple that can be answered without generating a network or any other techniques like data mining or statistical analysis, then it is not a wise decision to generate that network. Just for example, if we have a list of all students and their friends then we can't answer which student is more central by statistical analysis or data mining.

## Example

Consider we are interested in the Music domain and particularly interested in the study of different behaviors in the likeness of a song. We found that there are different websites offering services to users to listen and like songs online like SoundCloud. Can we extract information from SoundCloud? Yes, they offer an API to do so.

Who is a node? And what is an edge? One possible arrangement could be: a user is a node and an edge between two users exists if there is at least one song which is liked by both users. Or a node can represent a song and an edge if a song is liked by two users (or a specific number of users). You can observe that there can

be numerous kind of network we can generate and each network will represent certain kind of information.

<https://developers.soundcloud.com/docs/api/guide>

## Project Ideas

For this course, I have set domain to Pakistan – I encourage you to extract networks which are relevant to Pakistan. Some of example network are described below.

### Pakistan Highway Network

You have to generate and study the network of highway of Pakistan. There is a map available on internet (URL is given below) which shows highways with all major cities they are passing through. Rest you have generate network manually from this image. However, you might be able to get similar information online in digital format.

<http://nha.gov.pk/wp-content/uploads/2013/01/NHA-Road-Network-Maps-of-Projects-02.01.pdf>

### Pakistan's Airport Flight Map

There are number of airport in Pakistan– can we generate a network that represent the which airport has flight to which airport? Or something similar? You will have to extract information from websites of different airlines or flight schedules of different airport. Remember, we restricted to only airports of Pakistan.

### Railway Network

There are number of stations and there are certain routes which exists between these stations. One can extract information by exploiting Search feature of Pakistan Railway's website.

### Earthquake Network

US Geological Survey (USGS) provide the data of recent earthquake in different countries. The earthquake at one place can trigger earthquake at other place? We can create a relationship between two locations if they face earthquake within 5 minutes – or something else?

Data provided by USGS consist geo information as well. This means that one can draw the network on a world map as well.

<http://earthquake.usgs.gov/>

### Bus Routes

We have discussed Railway network above. Similarly, one can generate network for Bus routes. There are number of sites providing bus routes of different cities. Your task is to convert this information into network.

These all are separate projects

### Karachi Bus Routes

<http://www.travel-culture.com/pakistan/bus-routes-in-karachi.shtml>

<http://karachiroutes.blogspot.com/2012/10/karachi-bus-routes.html>

<http://www.kmc.gos.pk/Contents.aspx?id=12>

<http://www.urckarachi.org/bus%20routes%20karachi.pdf>  
[http://adiekhan.blogspot.com/2009/01/karachi-bus-routes\\_24.html](http://adiekhan.blogspot.com/2009/01/karachi-bus-routes_24.html)

**Lahore Bus Routes**

[https://en.wikipedia.org/wiki/List\\_of\\_bus\\_routes\\_in\\_Lahore](https://en.wikipedia.org/wiki/List_of_bus_routes_in_Lahore)

**Pakistan Inter-city bus routes (Daewoo)**

<http://www.daewoo.com.pk/schedules02.asp>

Many more ...