Assignment 9: Due Friday 7th June at 5pm

Late assignments will not be accepted except by prior arrangement (for a good reason)

Please include your student number in your handed up work, as Canvas doesn't give this to me automatically.

Collect a data set or choose a network dataset from a public data source; analyse the data; and derived a suitable model for the network.

- Everyone should choose a unique dataset.
- You must choose a dataset different from that in your honours/M.Phil project if you are already doing network analysis.
- You should choose something non-trivial.

Let me know your choice of data by Monday the 20th of September. **Marks will be deducted if you don't meet this deadline!!!**

Examples of possible sources:

- Biological General Repository for Interaction Datasets,
  `http://thebiogrid.org/`
- The House of Graphs
  `http://hog.grinvin.org/`
- IAM Graph Database Repository
  `http://www.iam.unibe.ch/fki/databases/iam-graph-database`
- Internet Topology Zoo
  `http://www.topology-zoo.org/index.html`
- Stanford GraphBase and Network Dataset Collection
  `http://people.sc.fsu.edu/~jburkardt/datasets/sgb/sgb.html`
  `https://snap.stanford.edu/data/index.html`
- Koblenz Network Collection (KONECT)
  `http://konect.uni-koblenz.de/networks/`
- Network of characters
  `https://figshare.com/articles/TV_Series_Networks_of_characters/2199646`
- Bitcoin network dataset
  `https://senseable2015-6.mit.edu/bitcoin/`
- Australian government data
  `https://data.gov.au/`
  `https://data.sa.gov.au/data/dataset`
- Data is beautiful at Reddit, e.g.,
  Minecraft `https://www.reddit.com/r/dataisbeautiful/comments/au6saj/minecraft_crafting_ingredients_network_oc/?st=k0ltj11w&sh=7e33e268`

Starcraft `https://www.reddit.com/r/dataisbeautiful/comments/adhuio/oc_interactive_starcraft_ii_player_network/?st=k0ltnj1f&sh=cef640d8`
and The Human Disease Network Graph, etc., etc.

but you are not limited to these.

Write a 5-6 page report detailing your dataset, its properties, and your model. Your report should be formatted in the style of the journal *IEEE Transactions on Network Science and Engineering*.

Hints and tips

1. Make sure you understand the network that is being modelled, and how it measured. Display this understanding by discussing
   - the type of network with reference to the taxonomy described in early lectures;
   - the precise nature of the nodes and edges;
   - the measurement methodology, its assumptions, and its limitations; and
   - any sampling that has been explicitly, or implicitly applied.

   You may need to be selective about datasets in order to ensure that you can find this information.

2. Modelling:
   - Don't limit yourself only to models I have described; and
   - Think about how are you going to convince me your model is the "right" model?
     - Think about metrics.
     - Look at properties described in modelling lecture.
     - Think about how the model might be used in an application.
   - But note that your model doesn't have to be great as long as you are critical about it, and clear about why you chose it.
   - Comparisons between alternatives will be looked on favourably.

3. Tools: there are lots of network analysis tools out there
   - igraph (R and Python)
   - NetworkX (Python)
   - Gephi (independent package)
   - Cytoscape (independent package)

   But there are lots of alternatives, including writing your own code.