Statistical Properties of Graphs

Network Science '21: Assignment 2

Prof. Dr Claudio J. Tessone, Dr Carlo Campajola

Blockchain & Distributed Ledger Technologies Group

A02.1 Average degree of the nearest neighbours

A02.1 Average degree of the nearest neighbours

Task: *For each dataset $\mathcal{G}$, randomise it to obtain a network $\mathcal{G}_{rnd}$ and explore their assortativity properties*

1. Plot the average degree of the nearest neighbours $k_{nn}(k)$ as a function of the vertices degree k
2. Compute the assortativity coefficient of the real network
3. Compute the assortativity coefficient of the randomised network

A02.2 Clustering and randomisation

A02.2 Clustering and randomisation

For each dataset $\mathcal{G}$, randomise it to obtain a network $\mathcal{G}_{rnd}$ and explore their clustering properties

1. Plot the degree distribution in double-logscale and mark with a vertical line the average degree $\langle k \rangle$
2. For each node, compute the clustering coefficient in the graphs $\mathcal{G}$ (denoted by $C(i)$) and $\mathcal{G}_{rnd}$ (denoted as $C_{rnd}(i)$)
3. Do a scatter plot of $C(i)$ vs. $C_{rnd}(i)$

A02.2 Notes on randomisation

- The randomised networks are obtained via multiple edge swaps via the `networkx` function `nx.algorithms.smallworld.random_reference`
- Make sure to set the parameter `connectivity = False` to have faster execution
- If execution is still slow, we provide randomised datasets
- We will see next that this is a simple way to randomise a network by preserving its degree distribution (can you see why this is the case?)
- This function does not ensure that the resulting graph is still connected (more on this later...)

A02.1 Datasets provided

Datasets provided:

- C. elegans interactomes: Nodes represent proteins and Edges represent protein-protein interactions in Caenorhabditis elegans (nematode) **[1]**
- AstroPhysics Arxiv collaborations: Nodes represent authors of papers submitted to arxiv.org in the AstroPh category and Edges represent co-authorship between two authors **[2]**
- Condensed Matter Arxiv collaborations: Nodes represent authors of papers submitted to arxiv.org in the CondMat category and Edges represent co-authorship between two authors **[3]**

A02.1 Datasets provided

- Kaggle chess players: Nodes represent chess players and Edges represent chess match among the world's top chess players **[4]**
- Dolphin social network: Nodes represent dolphins and Edges represent frequent associations observed among a group of 62 individuals **[5]**
- European airline network: Nodes represent airlines and Edges represent airline routes among European airports **[6]**

A02.1 Datasets provided

- Facebook friendships: Nodes represent Facebook users and Edges represent their friendship relations collected from survey participants **[7]**
- Florentine families: Nodes represent Florentine families during the Italian Renaissance and Edges represent marriage alliances and business relationships between two families **[8]**
- Game of Thrones coappearances: Nodes represent Game of Thrones characters and Edges represent coappearances of characters in the Game of Thrones series **[9]**

A02.1 Datasets provided

- Internet AS graph: Nodes represent autonomous systems and Edges represent connections between them **[10]**
- Jazz collaboration network: Nodes represent jazz musicians and Edges represent collaborations in bands that performed between 1912 and 1940 **[11]**
- 9-11 terrorist network: Nodes represent individuals and Edges represent their known social associations, centered around the hijackers that carried out the September 11th, 2001 terrorist attacks **[12]**

A02.1 Datasets provided

[1] N. Simonis et al., ”Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network.” Nature Methods 6(1), 47-54 (2009)
[2] Leskovec, J. Kleinberg and C. Faloutsos. ”Graph Evolution: Densification and Shrinking Diameters.” ACM Transactions on Knowledge Discovery from Data 1(1), article 2 (2007)
[3] Leskovec, J. Kleinberg and C. Faloutsos. ”Graph Evolution: Densification and Shrinking Diameters.” ACM Transactions on Knowledge Discovery from Data 1(1), article 2 (2007)

A02.2 Datasets provided

[4] Kaggle, ”Chess ratings - Elo versus the Rest of the World.” https://www.kaggle.com/c/chess/data (2010)
[5] D. Lusseau et al., ”The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations.” Behavioral Ecology and Sociobiology 54(4), 396-405 (2003)
[6] Alessio Cardillo, Jesús Gómez-Gardenes, Massimiliano Zanin, Miguel Romance, David Papo, Francisco del Pozo and Stefano Boccaletti. ”Emergence of network features from multiplexity.” Scientific Reports 3, Article number: 1344 doi:10.1038/srep01344

A02.2 Datasets provided

[7] J. Leskovec and J. J. Mcauley, Learning to discover social circles in ego networks, in Advances in Neural Information Processing Systems, 2012, pp. 539– 547.
[8] JF Padgett and CK Ansell, "Robust Action and the Rise of the Medici, 1400-1434". American journal of sociology, 1259-1319 (1993)
[9] A. Beveridge and J. Shan, "Network of Thrones." Math Horizons 23(4), 18-22 (2016)

A02.2 Datasets provided

[10] B. Karrer, M.E.J. Newman, and L. Zdeborova, ”Percolation on sparse networks.” Phys. Rev. Lett. 113, 208702 (2014)
[11] P. Gleiser and L. Danon, ”Community Structure in Jazz.” Advances in Complex Systems 6(4), 565-573 (2003)
[12] V. Krebs, ”Mapping networks of terrorist cells.” Connections 24, 43-52 (2002)

Prof. Dr Claudio J. Tessone, Dr Carlo Campajola

Blockchain & Distributed Ledger Technologies Group

tessone@ifi.uzh.ch

http://www.blockchain.uzh.ch