**DOCTOR OF PHILOSOPHY**

# A network theoretic approach to comparative mythology

MacCarron, Padraig

*Award date:*
2014

*Awarding institution:*
Coventry University

# A Network Theoretic Approach to Comparative Mythology

Pádraig Mac Carron

A thesis submitted for the award of
Doctor of Philosophy

May 2014

Applied Mathematics Research Centre, Coventry University

Supervisor: Prof. Ralph Kenna

# ABSTRACT

In recent years complex networks have been used to model a plethora of phenomena, many of which reside outside the realm of traditional sciences. Of these, mythology is one of the furthest removed fields from science. Many approaches to the study of comparative mythology exist and almost all are entirely qualitative. In this work, however, methods of network theory are applied to the myths and tales of different cultures in order to quantitatively compare them to one another.

In total, 33 mythological sources are analysed here. Social networks are constructed based on characters' interactions within each myth. The network properties allow us to make distinctions between the type of myth and, in some cases, to distinguish the myths of one culture from another. This method provides an entire new branch to the field of comparative mythology.

DECLARATION BY CANDIDATE

I hereby declare that this thesis is my own work and effort and that it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Signature: ..........................................................................................................

Date: ..................................................................................................................

## ACKNOWLEDGEMENTS

The author would like to thank the following people, without whom, this work could never have come to fruition.

Firstly I wish to thank my supervisory team, each of whom consistently made time for me no matter how busy they were. In particular I owe my deepest gratitude to my first supervisor, and the father of *mythematics*, Prof. Ralph Kenna. His boundless enthusiasm and constant encouragement continually inspired and motivated me. He has provided me with a vast array of opportunities and played an instrumental role in both my development as a researcher and a person. I am also indebted to Dr Thierry Platini, particularly for his continual patience when dealing with the mathematics behind many of the quantities in this work, and Dr Christian von Ferber, who on many late evenings at the office answered programming related questions.

I am deeply grateful to Cathal O'Conchobhair for both his friendship and his minutely thorough proof-reading of this work, and for adding lots of commas.

I wish to express my sincere thanks to my parents, Margaret and Pat, and my brother and sister, Brendan and Máirín, for their continual support and belief in me – not just over the past few years. Also to our dog, Rí-Rá, who continually teaches me lessons on life.

It has been an absolute pleasure to be a part of the Applied Mathematics Research Centre in Coventry University. I thank all the staff here, who always treated me as an equal, and many of whom have directly helped with this research such as Dr Rob Low, Dr Martin Weigel and Dr Taras Yavors'kii. I also thank the PhD students, interns and post-docs who I have shared an office with over the past three and a half years; namely Agnieszka, Antoine, Emilio, Eren, Franck, Hamid, Jon, Joseph, Kacper, Kélig, Lee, Lintao, Nathaniel, Richard, Robin, Simon and Taras. Thanks for putting up with me, in some cases keeping me sane, and in others trying to do the opposite!

I am grateful to all the contacts I have made from outside my own university while studying here in Coventry, in particular Sílvio Dahmen, Martine Barons, Yurij Holovatch, Olesya Mryglod and Johannes Zierenberg.

I thank all of my friends who have supported and encouraged me, far far too many to name individually, but most recently my housemate (and landlord) Wilfred Denga, and most consistently Thomas McSweeney.

Finally I wish to thank those involved with LAS Theatre for help bringing this research to the public, and the developers of the scientific libraries for Python; NumPy, SciPy, Matplotlib and NetworkX, all of which were used extensively throughout this work.

CONTENTS

# 1. INTRODUCTION

In recent years *complex systems* have been used to model a plethora of phenomena, many of which reside outside the realm of traditional sciences (Newman 2011). The overall behaviour of these systems tends to be extremely sensitive to a small variation in the initial conditions and fundamental laws are lacking. The collective behaviour exhibited is known as *emergence*: the properties of the resultant system are often more than just simple aggregates of the behaviours of its individual constituents.

Complex systems have been applied to a diverse array of fields, ranging from archaeology (Knappett *et al.* 2008) to zoology (Lusseau 2003), and many new mathematical methods have been developed to understand them. A particularly popular tool that has recently emerged is that of *complex network* theory (Watts & Strogatz 1998). Networks model the system as a set of interacting objects. In the work presented here, the techniques of network theory are applied to the study of mythology.

A mythology is a collection of myths generally belonging to a particular culture or religion. They are often used to support or justify a particular culture's rituals, beliefs and ethics (Leeming 2005). Almost every culture has some type of myth or belief system associated with it. A mythology is derived from a whole culture and not just an individual. In this sense, it can be seen as an *emergent property* of a society.

By applying the methods of complex networks to the social interactions of characters in mythological narratives, mythology is quantitatively explored. This allows for the properties of myths to be statistically compared and classified. This approach has already been recognised as opening up a new field of inquiry (Hazelkorn *et al.* 2013).

## *Sociophysics*

One of the first applications of methods from statistical physics to social interactions came from Serge Galam in a field he christened *sociophysics* (Galam *et al.*

1982). Numerous Galam theories employ the Ising model to predict the collective behaviour of groups of humans (Galam 2008). This is a model of a ferromagnetic spin system arranged on a lattice in which each spin interacts with each of its neighbours. While an individual spin has no direct influence on more distant spins, a re-orientation of a spin can influence its neighbours causing a disturbance to propagate over a large region of the lattice. This same principle is often used in the modelling of opinion dynamics and other collective phenomena (Galam 2008; Sznajd-Weron 2005).

More recently, a popular approach to sociophysics is to use *graphs* or complex networks to model social phenomena. The application of graph theory to study social relations and the nature of human interactions has a long history. Here we take this field a step further and study the societies portrayed in mythological narratives rather than modern societies.

### *Graph Theory and Social Networks*

The origins of *graph theory* can be traced back to a paper by Leonhard Euler in 1736 on the seven bridges of Königsberg. The question was asked as to whether one could find a path through the city of Königsberg which traversed each of its bridges just once. Euler proved that this is not possible and his study laid the foundations for graph theory.

Early applications of graph theory outside of mathematics include to chemical compositions and circuit diagrams. It was not until the 20th century that the first application of graph theory to social systems appeared. Graphs representing people and their interactions are usually referred to as *social networks*. The earliest example of such is usually attributed to Jacob Moreno (Moreno 1934). Moreno was interested in the social interactions between groups of people. In 1933, he produced the first social network diagram representing friendships between school children.

Interest in social networks quickly developed in behavioural science, sociology and anthropology as well as statistics and mathematics. Many statistics and quantities were developed in order to characterise properties of social networks. The concept of *small world* was formulated in the 1950s by mathematician Manfred Kochen and political scientist Ithiel de Sola Pool (see Schnettler (2009)). They based the idea on the anecdote of two strangers meeting and discovering they had a shared acquaintance. They presented the problem of determining the number of steps that chains of contacts spread in the population.

The psychologist Stanley Milgram took up the challenge of measuring the average number of steps separating two people. He devised a method to trace acquaintance chains by randomly allocating participants' letters and asking them to send them to preselected target persons. Participants could only send letters to someone that they knew on a first name basis. The average number of steps between the participant and target came to around six (Milgram 1967). This led to the the expression *six degrees of separation*, a phrase popularised by the John Guare play of that name.

In spite of global concepts such as small world however, early social network analyses were predominantly focused on measures of centrality and so-called *ego-centred* networks. Studies were focused more on the microscopic rather than macroscopic and large datasets were rare. In more recent years, social networks have been studied and analysed as part of the field of complexity science. Here, more interest is placed on the topological properties of the system rather than the individual.

### *Complex Networks*

The Galam approach to sociophysics often modelled human interactions using the Ising or Potts model in which each individual has the same number of neighbours. In a complex network however, connections are neither regular nor random. Complex networks allow for a different approach to the analysis of the structure of human relationships and a glimpse at their global emergent properties.

In their seminal paper, Watts & Strogatz (1998) observed similarities between the structure of the neural network of a nematode worm, a movie actor social network and the structure of a power grid. Since then a whole host of systems have been modelled and analysed using complex networks. Costa *et al.* (2011) detail a survey of applications ranging from technological networks to the US stock market. Core network properties allow for the comparison of the structures of such complex networks to one another.

While complex networks are ubiquitous in society, here we are interested in the properties of social networks in particular. Online resources such as the physics pre-print *arXiv* and the internet movie database *IMDb*, as well as online social networks, allow us to gather data on a much larger scale than previously available. Newman & Park (2003) demonstrate that social networks have different properties to those of other complex networks. Amaral *et al.* (2000) discriminate between different types of social networks, claiming that as the network of movie actors is economically driven, this is an inherently different system to the structure of school friendships. Gleiser (2007) even identifies differences between the social network of *Marvel Universe*'s comic book characters to more real-world social networks.

All of the above examples of social networks, however, are from the 20th century onwards. Applying these techniques to the characters and interactions that appear in myths may provide insights into the nature of human interactions for different, and less contemporary, cultures.

### *Mythological Networks*

There are many approaches to the study of comparative mythology. Some of these look for meaning in myth to explain natural phenomena (e.g. Barber & Barber (2006)), others search for common motifs and *universalities* of the human condition (Campbell 1949). However, almost all approaches thus far are qualitative. As myths tend to contain an abundance of characters, they provide a unique source for the construction of social networks. The analysis of their network properties allows for a first approach to quantitative approach comparative mythology.

In the field of folklore, the Aarne-Thompson classification catalogues thousands of folktales according to their plots and themes (Aarne & Thompson 1961). No similar system exists in the field of comparative mythology (Lyle 2006). Measuring the network properties which emerge from different narratives may allow us to find classifications for myths. It may also provide an insight into the similarities – or differences – between distinct cultures. Similarly, the methods of Gleiser (2007) may allow us to discriminate between fictional, superhero-like, myths and ones that are centred on a real society.

The following chapter introduces all the quantities from network theory that are used and measured in this thesis. In Chapter 3, these quantities are determined for readily available social network datasets in order to calibrate our approach and provide detailed empirical results for comparative purposes. In Chapter 4 the mythological data used and the methods employed to create the networks are discussed. In Chapter 5, the analyses of the data relating to various mythological narratives are presented. Chapter 6 presents an overview of the results and conclusions are drawn in the final chapter.

# 2. NETWORK THEORY

## *2.1 Networks*

A *graph* is an ordered pair $G = (\mathcal{V}, \mathcal{E})$ containing a set of vertices or nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. The number of vertices in the graph $N$ is given by $|\mathcal{V}|$ and the number of edges $L$ is given by $|\mathcal{E}|$. Each edge is a two-element subset of $\mathcal{V}$. A graph in which each pair of edges are in a specific order is called a *directed graph* or a *digraph*.

A graph can be represented by its *adjacency matrix* $\mathbf{A}$. This is an $N \times N$ matrix where each element $A_{ij} = 1$ if there is an edge between vertices $i$ and $j$, and $A_{ij} = 0$ otherwise. For an undirected graph $A_{ij} = A_{ji}$. If there are no self-loops (i.e. a vertex having an edge with itself) then $\mathrm{Tr}\mathbf{A} = 0$.

The most fundamental property of a vertex is its *degree*. The degree $k_i$ of vertex $i$ is defined by $k_i = \sum_j A_{ij}$. This is the number of edges connected to vertex $i$. The mean degree over the $N$ vertices of a network is given by

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i. \tag{2.1}$$

For an undirected graph, this can be expressed as

$$\langle k \rangle = \frac{2L}{N}. \tag{2.2}$$

An edge in a graph can also have a *weight*. This indicates its relative strength and can be represented by assigning a number to that edge. A graph with weighted edges is usually called a *network*.

From here on, the terms 'network' and 'graph' will be used interchangeably (as common in the literature). The networks contained herein are undirected and contain no self loops. The weight of an edge does not contribute to the degree of a vertex, therefore the degree refers to the number of neighbours associated with that vertex.

## 2.2 Degree Distribution

The distribution of the degrees in a network gives some important characteristics related to its structure. The *degree distribution* $p_k$ is the fraction of vertices in a network with degree $k$, [1]

$$p_k = \frac{1}{N} \sum_{i=1}^{N} \delta(k_i - k), \tag{2.3}$$

where $\delta(k_i - k)$ is the Kronecker delta function which, in this case, is 1 if $k_i = k$ and 0 if $k_i \neq k$.

The generating function associated with the degree distribution is given by

$$g(z) = \sum_k z^k p_k = \frac{1}{N} \sum_i z^{k_i} \tag{2.4}$$

The first moment of the degree distribution $\langle k \rangle$ is the mean of that distribution,

$$\langle k \rangle = \sum_k k p_k = g'(z)\big|_{z=1}. \tag{2.5}$$

The second moment

$$\langle k^2 \rangle = \sum_k k^2 p_k = g''(z)\big|_{z=1} + g'(z)\big|_{z=1} \tag{2.6}$$

is related to the variance $\sigma^2$ of the distribution by

$$\sigma_k^2 = \langle k^2 \rangle - \langle k \rangle^2. \tag{2.7}$$

The variance itself gives a measure of the spread of degrees about the mean.

### 2.2.1 Degree Distribution for Random Graphs

A *random graph* is a graph in which edges are distributed between pairs of vertices at random. One of the most common models is the Erdős-Rényi random graph (Erdős & Rényi 1959). Here an edge is created between each pair of vertices with probability $p$. A given vertex has $k$ edges to a maximum of $N - 1$. The probability for a vertex to have degree $k$ is proportional to $p^k(1-p)^{N-1-k}$ where the second term in the product represents the probability not to be linked to the

[1] In terms of notation, $p_k$ is used for discrete distributions and $p(k)$ for continuous distributions.

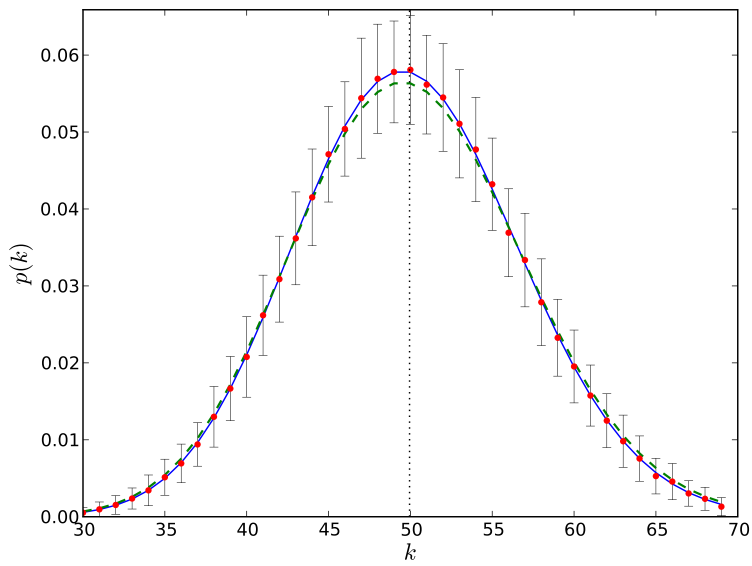



*Fig. 2.1:* The average degree distribution for 100 Erdős-Rényi random graphs each with $N = 1000$ and $p = 0.05$. Each point represents the average fraction of vertices with degree $k$ and the error bar is the standard deviation about that point. The continuous blue line is a binomial distribution (2.8) and the dashed green line is the Poisson distribution (2.14) The dotted vertical line represents the average degree of network.

$N - 1 - k$ other vertices. The degree distribution is therefore binomial

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (2.8)$$

Its corresponding generating function is

$$g(z) = [1 + p(z-1)]^{N-1}. \quad (2.9)$$

The first and second moments are then

$$\langle k \rangle = p(N-1), \quad (2.10)$$

$$\langle k^2 \rangle = p(N-1)[p(N-2)+1]. \quad (2.11)$$

The average degree distribution for 100 random graphs with $N = 1000$ and $p = 0.05$ is plotted in fig. 2.1. The average degree is $\langle k \rangle = 49.95$ and is marked by the vertical black dotted line. Eq. (2.8) is represented by the continuous blue line.

In the limit of large $N$, eq. (2.8) can be approximated by a Poisson distribution. For $N \gg k \geq 1$, Stirling's approximation $n! \approx \sqrt{2\pi}n^{n+1/2}e^{-n}$ can be applied to the first term in eq. (2.8) giving

$$\frac{(N-1)!}{k!(N-1-k)!} \approx \frac{N^{N+1/2}e^{N}}{k!(N-k)^{N-k+1/2}e^{N-k}} \approx \frac{1}{k!}N^k. \tag{2.12}$$

Assuming $p \ll 1$, then $\ln(1-p) \approx -p$, so the last term in eq. (2.8) can be approximated by

$$(1-p)^{N-1-k} = e^{(N-1-k)\ln(1-p)} \approx e^{-Np}. \tag{2.13}$$

Combining these, eq. (2.8) becomes

$$p_k \approx \frac{(Np)^k}{k!}e^{-Np}, \tag{2.14}$$

which is a Poisson distribution with mean $Np$. This is represented by the green dashed line in fig. 2.1.

### 2.2.2 Degree Distributions for Complex Networks

For complex networks, degree distributions are often found to have positive or right skew (Newman 2003b). These heavy tails are frequently due to a small number of vertices having very large degrees. Highly connected vertices are often known as *hubs*. In contrast, random graphs tend to have a lack of hubs due to having a small standard deviation in degree compared to the mean degree.

Fig. 2.2 (a) depicts the degree distribution for protein interactions in yeast (Jeong *et al.* 2001). The tail of this distribution is very noisy. This is because there are multiple instances of only one vertex with a specific degree, e.g. more than one vertex for which $p_k = 1/N$. A common method to deal with this is to consider the complementary cumulative distribution function (Newman 2005). This is the probability a vertex has degree greater than $k$,

$$P_k = \sum_{j=0}^{\infty} p_{k+j} = \sum_{q=k}^{\infty} p_q. \tag{2.15}$$

The complementary cumulative distribution function for protein interactions in yeast is shown in fig. 2.2 (b).

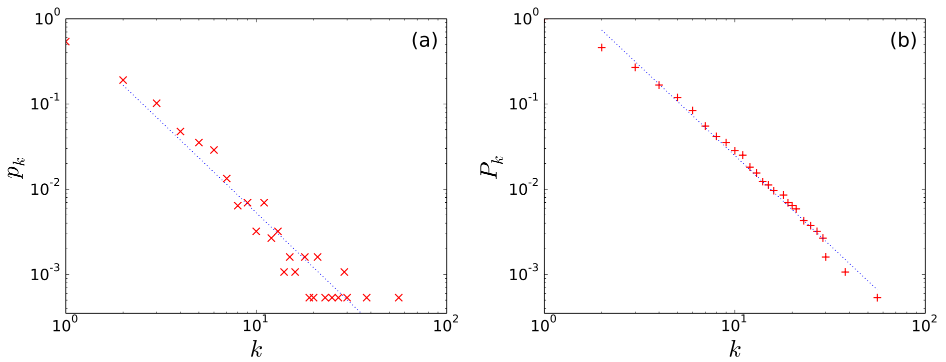



*Fig. 2.2:* (a) The degree distribution on a log-log scale for protein interaction network in yeast (Jeong *et al.* 2001) with a fitted power law, observe that the tail is very noisy. (b) The complementary cumulative distribution function for the same network, again with a fitted power law.

Amaral *et al.* (2000) find three classes of degree distribution for small-world networks (see section 2.6 for details on small-world networks). These are (i) power-law degree distributions, (ii) truncated power-law distributions which are characterised by power-law regimes followed by a sharp cut-off, and (iii) degree distributions with fast decaying tails such as exponentials or Gaussians. We discuss these in turn.

### (i) *Power-Law Distributions*

A distribution commonly observed in complex networks is that described by a power law (eg. Strogatz (2001); Albert & Barabási (2002)). This has the form

$$p_k \sim k^{-\gamma}, \tag{2.16}$$

where $\gamma \geq 1$. Usually the power-law behaviour is found only in the tail of the distribution beginning at some value of degree $k = k_{\text{min}} > 0$. Jeong *et al.* (2001) show that the degree distribution for the protein interactions in yeast is fitted by a power law. This is represented by the dotted line in fig. 2.2 (a) with $k_{\text{min}} = 2$ .

A network with a power-law distribution is often called a *scale-free* network. If the scale of the degree $k$ is increased by some factor $a$, the shape of the distribution is unchanged except for an overall multiplicative constant (Newman 2005),

$$p(ak) = a^{-\gamma} p(k) \sim p(k). \tag{2.17}$$

However, while the degree distribution may indeed be scale free, there are usually scales present in other network properties.

Albert & Barabási (2002) find that complex networks often have exponents in the range $2 \leq \gamma \leq 3$. If, for large $N$, we approximate the distribution as continuous $p(k)$ and take the second moment we obtain

$$\langle k^2 \rangle = \int_{k_{\min}}^{N-1} k^2 p(k) dk \sim \frac{1}{3-\gamma} \left[ k^{3-\gamma} \right]_{k_{\min}}^{N} . \tag{2.18}$$

As $N$ goes to infinity, one observes that the second moment diverges when $\gamma \leq 3$. Therefore for the networks presented here, if the degree distributions follow a power law with $\gamma \leq 3$, they would be expected to have a large value of $\langle k^2 \rangle$. In a similar manner, the first moment $\langle k \rangle$ diverges in an infinite network with $\gamma \leq 2$, however empirically exponents are not commonly found in this regime.

Remaining in the continuous regime, the complementary cumulative form of a power-law degree distribution is given by (see A.8)

$$P(k) \sim k^{1-\gamma}. \tag{2.19}$$

This preserves the power-law form but increases the value of the exponent. Hence a power law is also depicted in fig. 2.2 (b).

A two-parameter variation of the power law is sometimes fitted to degree distributions (e.g. Gleiser & Danon (2003)). This has the form

$$p_k \sim (1 + \alpha k)^{-\gamma} \tag{2.20}$$

for parameter $\alpha > 0$. This distribution is no longer scale-free but the first and second moment diverge in the same range as for the standard power law.

(ii) *Truncated Power-Law Distributions*

The second class of network described by Amaral *et al.* (2000) contains a power-law regime followed by a sharp cut-off. Albert & Barabási (2002) provide a list of exponents and corresponding cut-offs in a range of complex networks.

A power law with an exponential cut-off is known as a truncated power law and has the form

$$p_k \sim k^{-\gamma} e^{-k/\kappa}. \tag{2.21}$$

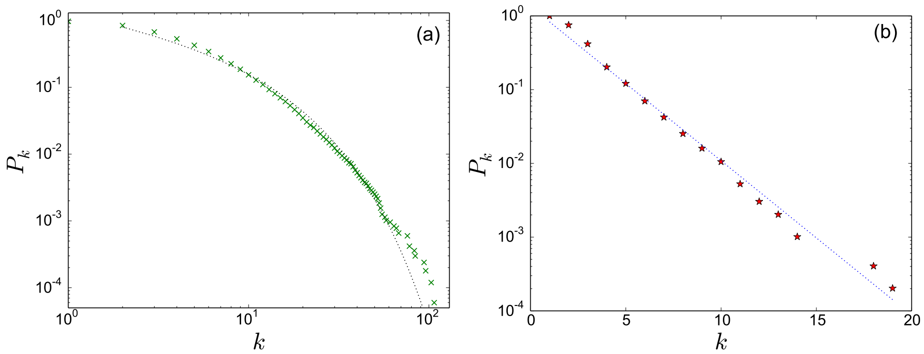



*Fig. 2.3:* (a) The complementary cumulative degree distribution on a log-log scale for the scientific co-authorship network of the cond-mat arXiv (Newman 2001b) with a fitted truncated power law. (b) The complementary cumulative degree distribution for a US power grid network (Watts & Strogatz 1998) on a log-linear scale with a fitted exponential.

A truncated power law is shown for the scientific collaboration network of the condensed matter arXiv from 1995 to 1999 in fig 2.3 (a) (Newman 2001b). This is represented by the black dotted line in fig. 2.3 (a).

(iii) *Exponential and Related Distributions*

The exponential distribution is another commonly encountered degree distribution in complex networks (Newman 2003b). It is given by

$$p_k \sim e^{-k/\kappa}, \tag{2.22}$$

where the parameter $\kappa > 0$ sets a scale.

Fig. 2.3 (b) depicts the degree distribution for the power-grid network of the western states of the US (Watts & Strogatz 1998). This is shown on a log-linear scale where the dotted blue is an exponential distribution.

Other distributions involving the exponential function commonly found in the study of complex networks (Amaral *et al.* 2000; Newman 2003b) and which are used here are:

- the Gaussian (or normal) distribution which has the form

$$p_k \sim \exp\left[-\frac{(k-\mu)^2}{2\sigma^2}\right], \tag{2.23}$$

where $\mu$ is the mean and $\sigma^2$ is the variance of the distribution.

- the log-normal distribution which is given by

$$p_k \sim \frac{1}{k} \exp\left[-\frac{(\ln k - \mu)^2}{2\sigma^2}\right]. \tag{2.24}$$

  This is more strongly skewed than the Gaussian distribution.

- the stretched exponential

$$p_k \sim e^{-(k/\kappa)^\beta}. \tag{2.25}$$

  If the parameter $\beta = 1$, then this is an exponential distribution. If $\beta > 1$ this is called the compressed exponential function. It recovers the Gaussian distribution when $\beta = 2$.

- the Weibull distribution

$$p_k \sim \left(\frac{k}{\kappa}\right)^{\beta-1} e^{-(k/\kappa)^\beta}. \tag{2.26}$$

  The stretched exponential is the complementary cumulative function of the Weibull distribution. For both the stretched exponential and the Weibull distribution, the concavity changes after the mean of the distribution when parameter $\beta$ goes from $\beta < 1$ to $\beta > 1$.

The method of maximum likelihoods will be used to estimate the parameters of these distributions. Details of the log-likelihood for each distribution are discussed in Appendix A.

## 2.3 Maximum Likelihood Estimators

As mentioned in section. 2.2, the complementary cumulative distribution function in eq. (2.15) is often used to reduce the noise in the tail of the probability distribution. Fits are then often made to the cumulative distribution $P_k$ rather than the original degree distribution $p_k$.

In recent years it has been suggested that applying the method of least squares to cumulative distribution is unsuitable for empirical data (Edwards *et al.* 2007; Clauset *et al.* 2009). This is in part due to the data being discrete and therefore the continuous approach to the complementary cumulative form may not be appropriate (in particular when dealing with small datasets). Further reason cited are the estimates for the parameters are subject to large errors and furthermore the errors are difficult to estimate (Clauset *et al.* 2009). Instead it is claimed that Maximum Likelihood Estimators (MLE) provide better measurements for the parameters as no estimator has lower asymptotic error in the limit of large sample size as the MLE (Clauset *et al.* 2009).

The *likelihood* $\mathcal{L}$ is the probability that an independent and identically distributed data-set of $N$ observations were drawn from a model $p(k)$ with parameter $\theta$,

$$\mathcal{L}(\theta|k) = \prod_{i=1}^{N} p_\theta(k_i). \tag{2.27}$$

The data are most likely to have been generated by the model with the value $\theta$ that maximises the likelihood. It is often easier to deal with the logarithm of the likelihood which has a maximum at the same parameter value.

To calculate the error $\sigma_\theta$ in the parameter $\theta$, we use the variance of the most likely value $\hat{\theta}$ when it is normally distributed (Fisher 1922),

$$\sigma_\theta^2 = -\frac{1}{N}\left\langle \frac{\partial^2 \log p(k|\theta)}{\partial \theta^2} \right\rangle^{-1}. \tag{2.28}$$

The expectation term is often called the Fisher information at $\theta$. Alternatively, an estimate for the error can be obtained by the method of 'bootstraping'.

Both MLEs or least squares can give reasonable estimates for the parameters of a given distribution, however this does not mean that the distribution is actually a good model for the data. To compare different models the *Akaike information criterion* for small sample size $\mathrm{AIC}_c$ (Akaike 1974; Burnham & Anderson 2002) or the *Bayesian information criterion* BIC (Schwarz 1978) may be employed. The $\mathrm{AIC}_c$ is given by

$$\mathrm{AIC}_c = -2\ln \mathcal{L}(\hat{\theta}|k_i) + 2n_\theta + \frac{2n_\theta(n_\theta+1)}{N-n_\theta-1}, \tag{2.29}$$

where $n_\theta$ is the number of parameters in the model. When comparing different models, the same numbers of data points must be used. Therefore, if fitting to the tail of a distribution, the same $k_{\min}$ must be chosen in each case.

The $\mathrm{AIC}_c$ weights $w_m$ of a particular model $m$ gives the relative likelihood of $R$ models as

$$w_m = \frac{e^{-\Delta_m/2}}{\sum_{r=1}^{R} e^{-\Delta_r/2}}, \tag{2.30}$$

where $\Delta_m = \mathrm{AIC}_{c_m} - \mathrm{AIC}_{c_{\min}}$ and $\mathrm{AIC}_{c_{\min}}$ is the minimum value of $\mathrm{AIC}_c$ for the $R$ models. The weight $w_m$ denotes the extent to which model $m$ is favoured. Note that this is not a goodness of fit test; it does not give an indication of how well a model fits the data. Rather, it signals the most likely of the candidate models used.

The BIC is given by

$$\mathrm{BIC} = -2 \ln \mathcal{L}(\hat{\theta}|k_i) + n_\theta \ln N. \tag{2.31}$$

A similar expression for the BIC weights as eq. (2.30) is used to determine the most likely candidate model.

Further details on the Information Criteria are provided in Appendix A.3.

## 2.4 Paths and Connectivity

A *path* in a graph is a sequences of edges between two vertices. If there exists a path between every pair of nodes in the graph then the graph is said to be *connected*. If a graph is not connected, the largest connected sub-component is called the *giant component*.

The shortest path, or *geodesic*, $\lambda_{ij}$ is the path that traverses the minimal number of edges between vertices $i$ and $j$. If there is no path between $i$ and $j$ then by convention $\lambda_{ij} = \infty$. The average path length $\ell$ is defined by summing over all finite paths of each component. Defining $\{\mathcal{C}_m\}$ as the set of components in $G$, the average path length is defined by

$$\ell = \sum_m \frac{1}{n_m(n_m - 1)} \sum_{i,j \in \mathcal{C}_m} \lambda_{ij}, \tag{2.32}$$

where $n_m$ denotes the number of nodes in each component. The longest geodesic $\ell_{\max}$ is known as the *diameter* of the network.

### 2.4.1 Average Path Length for Random Graphs

Molloy & Reed (1995) show that for the Erdős-Rényi random graph, if $\sum_k k(k-2)p_k > 0$ then the graph almost surely connected. This can be re-expressed as

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2, \tag{2.33}$$

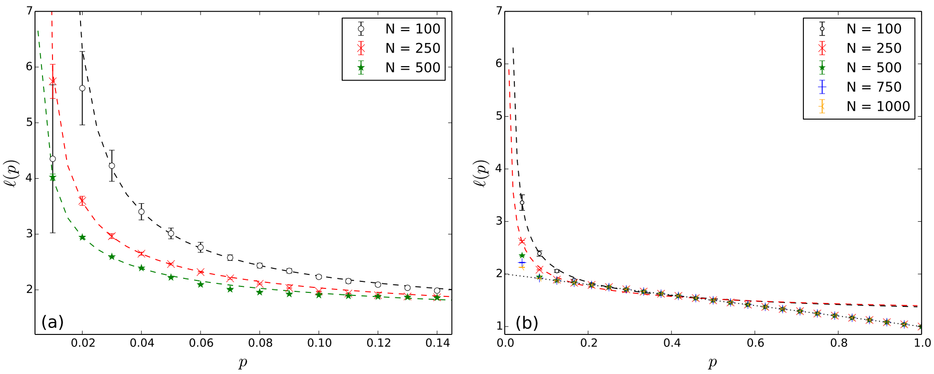



*Fig. 2.4:* The mean of the average path lengths $\ell$ for 50 random graphs with different system sizes versus the probability $p$ for creating edges. The error is given by the standard deviation of $\ell$. Panel (a) shows eq. (2.36) as the dashed line for three different system sizes. Panel (b) shows the entire range of $p$ however and as the probability becomes large, eq. (2.36) no longer fits. Above $p = 0.5 + a/\ln N$, the data is better fitted by the dotted line $\ell(p) = 2 - p$.

and using eq. (2.10) and eq. (2.11), this becomes

$$p > \frac{1}{N-2}. \tag{2.34}$$

Above the critical threshold $p_c = 1/(N-2)$ is where percolation occurs (Cohen *et al.* 2000).

Albert & Barabási (2002) find that the diameter for a random graph is approximately

$$\ell_{\max_{\mathrm{rand}}} = \frac{\log N}{\log\langle k\rangle}, \tag{2.35}$$

noting that for $N \gg 1$, $\langle k\rangle \approx pN$. They also suggest that the average path length $\ell_{\mathrm{rand}}$ scales with $\log N/\log\langle k\rangle$ in a proportional manner. However, a more accurate expression is given by (Fronczak *et al.* 2004)

$$\ell_{\mathrm{rand}} = \frac{\ln N - a}{\ln pN} + \frac{1}{2}, \tag{2.36}$$

where $a \simeq 0.5772$ is the Euler-Mascheroni constant.

In fig. 2.4 (a) the mean $\ell_{\mathrm{rand}}$ is shown for three different systems sizes with increasing probability for creating edges $p$. For each system size 50 random graphs were generated. The mean and standard deviation of $\ell_{\mathrm{rand}}$ are shown as the points and error-bars respectively. Eq. (2.36) is the fitted to each data set. In fig. 2.4 (b), however, $\ell_{\text{rand}}$ is plotted for the entire range of $p$. This shows that the behaviour becomes linear for large $p$ and has the form

$$\ell_{\text{rand}}(p) = 2 - p. \tag{2.37}$$

Substituting this into eq. (2.36), and assuming $\ln N \gg \ln p$, we find that

$$p \approx \frac{1}{2} + \frac{\gamma}{\ln N}. \tag{2.38}$$

Thus there are three regimes characterising the behaviour of $\ell_{\text{rand}}(p)$:

- $p < \frac{1}{N-2}$, this is before the percolation threshold. The graph is likely to be fragmented and contain many small components.
- $\frac{1}{N-2} < p < 0.5 + \frac{\gamma}{\ln N}$, there will almost surely be a fully connected component and $\ell_{\text{rand}}$ follows eq. (2.36).
- $p > 0.5 + \frac{\gamma}{\ln N}$, here $\ell_{\text{rand}}$ is very short ($\ell < 2$) and has the linear form of eq. (2.37).

Complex networks are generally found to have average path lengths comparable to the average path lengths of random graphs (Watts & Strogatz 1998).

## 2.5 Clustering Coefficient

Another measure of connectivity in a graph is the *clustering coefficient*. This measures the probability of two neighbours of a vertex sharing an edge. The clustering coefficient of vertex $i$ is given by

$$C_i = \frac{2n_i}{k_i(k_i - 1)}, \tag{2.39}$$

where $n_i$ is the number of edges linking the $k_i$ neighbours of vertex $i$ to each other (Watts & Strogatz 1998). The *mean clustering coefficient* $C$ of the entire network is obtained by averaging eq. (2.39) for all $N$ vertices,

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{2.40}$$

This is sometimes known as the Watts-Strogatz clustering coefficient. In the case of single degree vertices (*leaf* nodes) $k_i = 1$ where $n_i = k_i - 1 = 0$, we define

$C_i = 0$. Note that some authors instead define $C_i = 1$ for vertices with $k_i = 1$ (e.g. Brandes & Erlebach (2005)) and others ignore vertices with $k_i = 1$ entirely for the purposes of clustering coefficient calculations (eg. Latapy (2008)). These however give artificially large values for the clustering coefficient. For example in fig. 2.5, there are 7 vertices with only one closed triangle. The last two definitions both give a clustering coefficient of $C > 0.5$. (For a full discussion on the effect of leaf nodes on the clustering coefficient, see Kaiser (2008).)

An alternative measure for the clustering coefficient, sometimes known as the transitivity (and which will be referred to as such from here on to avoid confusion), is commonly used in the sociology literature (Wasserman 1994). Denoting $N_\Delta$ as the total number of triangles in the network and $N_t$ as the number of connected triplets (i.e. paths of length 2), then

$$C_T = \frac{3N_\Delta}{N_t}. \tag{2.41}$$

This is a global quantity as opposed to the average of the local quantities of eq. (2.39). The Watts-Strogatz version gives more weight to low degree vertices, as the denominator in eq. (2.39) is small for such vertices (Newman 2003b). An illustration outlining the difference is given in fig. 2.5. Empirically we find for social networks that the clustering coefficient $C$ is usually greater than the transitivity $C_T$.

### 2.5.1 Clustering Coefficient for Random Graphs

For a random Erdős-Rényi graph, the probability that any two vertices have an edge is the same regardless of whether they have a common neighbour, hence

$$C_{\text{rand}} = p = \frac{\langle k \rangle}{N-1}, \tag{2.42}$$

where eq. (2.10) has been used. For $N \gg \langle k \rangle \geq 1$, $C_{\text{rand}}$ tends to zero.

This is quite a stark contrast to complex networks which are often found to have high clustering coefficients despite having relatively small mean degrees (Albert & Barabási 2002). However, unlike the Erdős-Rényi graph, complex networks' degree distributions are not binomial. Attention is next turned to a random graph with an arbitrary degree distribution $p_k$ and the behaviour of the clustering coefficient is evaluated.

The probability for two vertices $i$ and $j$ to both be connected to vertex $v$ is

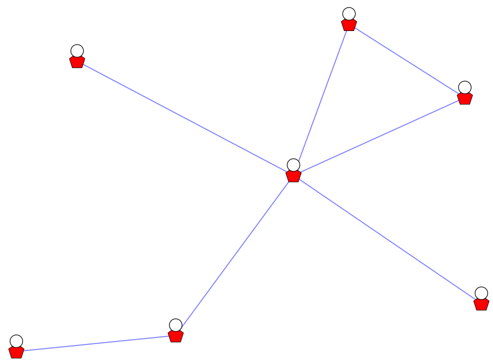


*Fig. 2.5:* Illustration of the clustering coefficient. There is only one closed triad of 13 connected triplets. This gives a value of $C_T = 3/13 \approx 0.23$. The Watts-Strogatz definition from eq. (2.40) gives $C = 0.3$. If instead the Watts-Strogatz clustering is only calculated for vertices with $k_i > 1$, then $C = 0.53$ and if for $k_i = 0$, $C_i = 1$ then $C = 0.73$. These values are quite large considering there is only one triangle. For a random graph of the same size and average degree, $C_{\text{rand}} = 0.29$.

given by

$$p_{v,ij} = \frac{(k_i - 1)(k_j - 1)}{2L}, \tag{2.43}$$

where it is assumed that the number of edges $L \gg 1$. The probability for any neighbour of $v$ to share an edge is then

$$n_v = \sum_{ij, i \neq j} A_{vi} A_{vj} p_{v,ij}. \tag{2.44}$$

For a graph generated at random, the probability to connect two vertices $i$ and $j$ is proportional to the product of their degrees $k_i k_j$. Hence the approximation

$$A_{ij} \approx \frac{k_i k_j}{2L} \tag{2.45}$$

can be made. Combining eq. (2.44) and eq. (2.45), then,

$$n_v \approx \frac{k_v^2}{(2L)^3} \sum_{ij} k_i k_j (k_i - 1)(k_j - 1)(1 - \delta_{ij}), \tag{2.46}$$

where the $1 - \delta_{ij}$ term ensures there are no loops. Evaluating the summation and neglecting $\mathcal{O}(1/N)$ terms, $n_v$ can be expressed as

$$n_v \approx \frac{k_v^2}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{2.47}$$

Returning to eq. (2.39), the clustering coefficient of $v$ is

$$C_v \approx \frac{1}{N} \frac{k_v}{k_v - 1} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{2.48}$$

For the entire network

$$C \approx \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3} \sum_v \frac{k_v}{k_v - 1}. \tag{2.49}$$

In section 2.2.2, it is shown that $\langle k^2 \rangle$ diverges for a power-law distribution when $\gamma \leq 3$. As the exponents in many complex networks are commonly found in the range $2 \leq \gamma \leq 3$ (Albert & Barabási 2002), the clustering coefficient is no longer dominated by factor of $N^{-1}$.

A naïve expression for the transitivity $C_T$ can also be obtained which we denote as $C_\mathrm{n}$. The number of triangles $N_\Delta$ can be estimated by

$$N_\Delta \approx a_1 \sum_v n_v = \langle k^2 \rangle \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}, \tag{2.50}$$

for some constant $a_1$. The number of connected triplets, $N_t$ is obtained from

$$N_t = a_2 \sum_i \sum_{j, j \neq i} \sum_{k, k \neq i, j} A_{ij} A_{jk}, \tag{2.51}$$

with some constant $a_2$. Employing eq. (2.45) and neglecting $\mathcal{O}(1/N)$ terms again, this becomes

$$N_t \approx a_2 N \langle k^2 \rangle. \tag{2.52}$$

Therefore eq. (2.41) can be expressed as

$$C_\mathrm{n} \approx \frac{a_1}{a_2} \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{2.53}$$

For the Erdős-Rényi model for large $N$, eq. (2.53) gives $C \approx p\, a_1/a_2$ which is

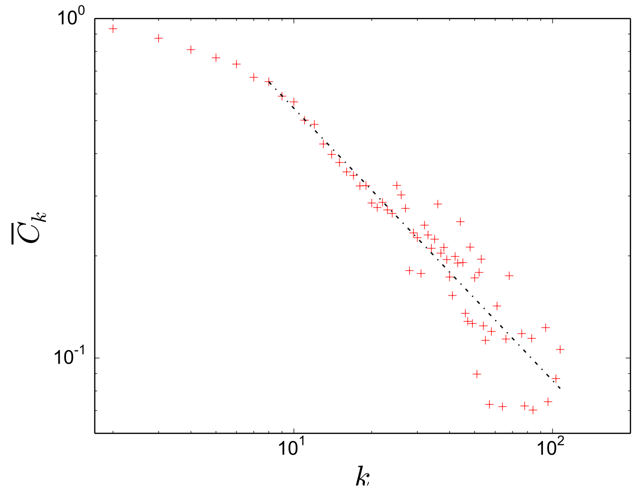



*Fig. 2.6:* Mean clustering coefficient $\bar{C}_k$ per degree for the cond-mat arXiv coauthorship network from 1995-1999 (Newman 2001b). The dashed black line is a power law of the form $\bar{C}(k) \sim k^{-0.80}$.

eq. (2.42) when $a_1 = a_2$. Therefore we assume

$$C_{\mathrm{n}} \approx \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}. \tag{2.54}$$

An alternate derivation of eq. (2.54) can be found in Newman & Park (2003).

### *2.5.2 Hierarchy*

It has also been suggested that if the clustering coefficient $C_i$ decreases as a power of the degree $k_i$, the network is *hierarchical* (Ravasz & Barabási 2003). As there can be multiple values of $C_k$ for degree $k$, the mean clustering coefficient per degree $\bar{C}_k$ is used,

$$\bar{C}_k \sim k^{-\beta}. \tag{2.55}$$

Ravasz & Barabási (2003) observe that the exponent is often found to be $\beta \approx 1$. In fig. 2.6, this is shown for the cond-mat arXiv coauthorship network from 1999-2003 (Newman 2001b) with a fitted power law with exponent $\beta = 0.80 \pm 0.04$.

In practice however, a power-law may not describe the data well (see Gleiser (2007); Luduena *et al.* (2013) for example). Nonetheless, a decay indicates that high degree vertices tend to have low clustering coefficients. In many sub-communities such nodes play an important role in keeping the entire network intact.

## 2.6 Small Worldness

A common property of complex networks is that most vertices can reach other vertices in just a small number of steps relative to the size of the network (Estrada 2011). This is known as the *small-world effect* (Milgram 1967).

A network is often said to be small world if it's average path length $\ell$ scales logarithmically with the size of the network $N$ (Barrat & Weigt 2000),

$$\ell \sim \ln N. \tag{2.56}$$

Here, however, we define the small worldness based on the Watts & Strogatz model (Watts & Strogatz 1998). A network is small world if it meets the following two criteria:

- The average path length is similar to the average path length of a random graph of the same size and average degree $\ell \approx \ell_{\mathrm{rand}}$.
- The clustering coefficient is much larger than the clustering coefficient of a random graph of the same size and average degree $C \gg C_{\mathrm{rand}}$.

A recent suggestion for a quantitative determination of small worldness is given by

$$S = \frac{C/C_{\mathrm{rand}}}{\ell/\ell_{\mathrm{rand}}}. \tag{2.57}$$

The network is then small world if $S > 1$ (Humphries & Gurney 2008).

## 2.7 Signed Graphs and Structural Balance

In addition to a weight, an edge in a graph can be assigned a positive or negative sign. A *signed graph* is a pair $(G, s)$ that consists of a graph $G$ and a sign mapping $s$ from $E$ to the sign group $\{+, -\}$. In a social network, these are used to distinguish between friendly and hostile interactions. *Friendly edges* are denoted as positive and *hostile edges* are negative. These are described in more detail in section 4.2.

In the overall network, closed triads with just one hostile edge are disfavoured as a single hostile link prompts the opposite node in a triangle to take sides. The propensity to disfavour odd numbers of hostile links in a closed triad is known as *structural balance* (Heider 1946; Cartwright & Harary 1956). This is related to the notion of "the enemy of my enemy is my friend."

Structural balance is normally used as a dynamical property. For example, it has been observed in the shifting alliances of nations in the lead-up to war (Antal *et al.* 2006). However, it has also been analysed statically by measuring the abundance of triangles with an odd number of positive edges. This method has been used in the social network of an online multiplayer game (Szell & Thurner 2010).

## 2.8 Assortativity

The tendency for individuals to associate with those who have similar attributes (e.g. ethnicity, religious beliefs, etc.) to themselves is known as *homophily* (McPherson *et al.* 2001). In complex networks, this is similar to the notion of *assortativity* (Newman 2002). Assortativity measures the correlations between properties of vertices and their neighbours.

The most common attribute studied is the *degree assortativity*. A network with correlations between the degrees of vertices is said to be assortatively mixed by degree, and a network with anti-correlations is disassortatively mixed by degree. In a random graph there are no correlations between the degrees of neighbouring vertices.

Consider the two vertices at the extremities of a randomly chosen edge in a graph. The probability of one vertex having degree $k$ and the other degree $q$ is

$$\Phi(k,q) = \frac{1}{2L}\sum_{i,j} A_{ij}\delta(k_i - k)\delta(k_j - q). \tag{2.58}$$

Summing over the $L$ edges, the average degree of the vertex at the end of an edge is denoted by $\mathrm{E}[k]$. As we are dealing with undirected graphs $\mathrm{E}[k] = \mathrm{E}[q]$. It is important to note that $\mathrm{E}[k]$ is not the same as the average degree of the network $\langle k \rangle$ (which sums over the $N$ vertices), these quantities are related however, as shown below. Summing over the degree $q$, the marginal probability is given by

$$\phi(k) = \sum_{q} \Phi(k,q). \tag{2.59}$$

From this $\mathrm{E}[k]$ is evaluated by

$$\mathrm{E}[k] = \sum_k k\phi(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}, \tag{2.60}$$

and we observe the relationship between the average degree at the end of an edge $\mathrm{E}[k]$ and the average degree of the network $\langle k \rangle$.

The quantity $\mathrm{E}[kq]$ is given by

$$\mathrm{E}[kq] = \sum_{k,q} kq\Phi(k,q). \tag{2.61}$$

If the graph is uncorrelated then $\Phi(k,q) = \phi(k)\phi(q)$. This implies,

$$\mathrm{E}[kq] = \mathrm{E}[k]\mathrm{E}[q]. \tag{2.62}$$

A network has positive correlations when $\mathrm{E}[kq] > \mathrm{E}[k]\mathrm{E}[q]$, and negative correlations when $\mathrm{E}[kq] < \mathrm{E}[k]\mathrm{E}[q]$. The degree assortativity $r_k$ is then defined as

$$r_k = \frac{1}{\sigma_k^2}(\mathrm{E}[kq] - \mathrm{E}[k]\mathrm{E}[q]), \tag{2.63}$$

where $\sigma_k^2$ is the variance of $k$

$$\sigma_k^2 = \mathrm{E}[k^2] - \mathrm{E}[k]^2. \tag{2.64}$$

The variance acts as a normalisation ensuring $-1 < r_k < 1$. If $r_k > 0$ the network is assortative, if $r_k < 0$ it is disassortative and if $r_k = 0$ there are no correlations between the degrees of vertices.

Using eq. (2.58), the assortativity in eq. (2.63) can be re-expressed as

$$r_k = \frac{\sum_{ij} A_{ij}(k_i - \mathrm{E}[k])(k_j - \mathrm{E}[k])}{\mathrm{E}[k^2] - \mathrm{E}[k]^2}, \tag{2.65}$$

which is the standard form of the Pearson correlation coefficient.

The expected statistical error $\sigma_k$ is calculated using the jackknife method (Efron 1979),

$$\sigma_r^2 = \sum_{e=1}^{L}(r_e - r)^2, \tag{2.66}$$

where $r_e$ is the value of the $r$ when the edge $e$ is removed (Newman 2003a).

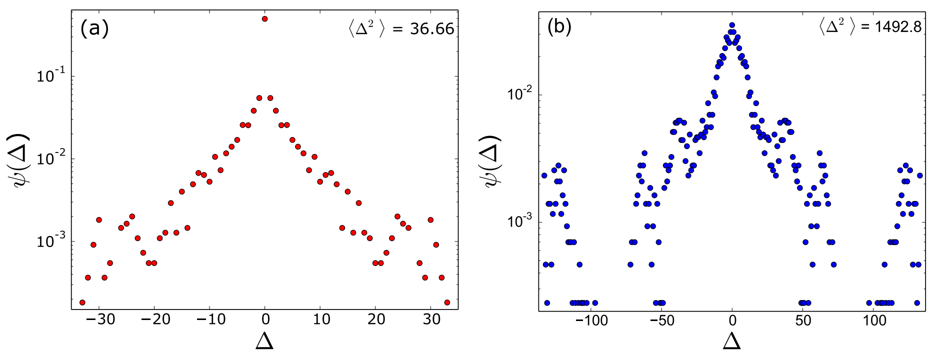



*Fig. 2.7:* The $\psi(\Delta)$ distribution for two networks. Panel (a) is the assortative network of coauthors on network science (Newman 2006). Panel (b) shows the disassortative of the neural network of a nematode worm (Watts & Strogatz 1998). The disassortative distribution has much larger and more frequent fluctuations in the tail.

### 2.8.1 Degree Difference in Edges

A method of visualising the assortativity introduced here is presented as follows. From eq. (2.65), one observes that a network is assortative if, on average, $k_i$ and $k_j$ are both above or are both below the mean $\mathrm{E}[k]$ for any connected pair of vertices $i, j$. We next define the probability $\Psi(K, \Delta)$ for the sum and the difference of the degree of each vertex to be equal to $K$ and $\Delta$ respectively. This is given by (Mac Carron *et al.* 2014)

$$\Psi(K, \Delta) = \frac{1}{2L} \sum_{i,j} A_{ij} \delta(k_i + k_j - K) \delta(k_i - k_j - \Delta). \tag{2.67}$$

The marginal probability $\psi(\Delta) = \sum_K \Psi(K, \Delta)$ gives the probability of finding two vertices (at the extremity of the same edge) such that $k_1 - k_2 = \Delta$. Plotting this distribution allows one to visualise the behaviour of the assortativity. Large fluctuations far from the mean $\langle \Delta \rangle$ cause a network to become more disassortative. This function is symmetric for undirected networks.

Fig. 2.7 (a) shows the $\psi(\Delta)$ distribution for the coauthorship network of researchers on network science on a log-linear scale (Newman 2006). This is an assortative network with $r_k = 0.46$. The distribution is uniformly decaying away from the mean $\langle \Delta \rangle = 0$. A large fraction of vertices that share edges also have the same degree. Fig. 2.7 (b) is the distribution for the disassortative network ($r_k = -0.16$) of neural network of the *C. elegans* nematode worm (Watts & Stro-

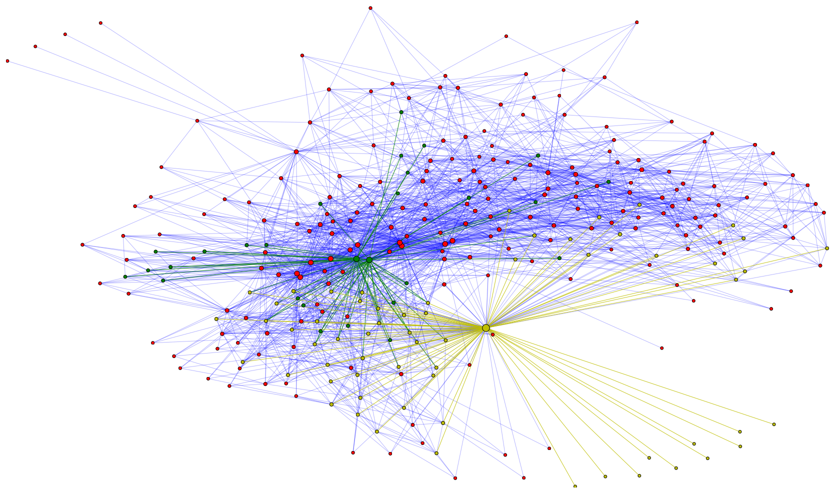


*Fig. 2.8:* The neural network of a nematode worm (Watts & Strogatz 1998). Edges and their correspond vertices that have a degree difference between $133 \leq |\Delta| \leq 122$ are coloured yellow, and edges and vertices with $66 \leq |\Delta| \leq 62$ are coloured green. The average degree of vertices at the end of an edges is $\mathrm{E}[k] = 26.05$.

gatz 1998; White *et al.* 1986). Here we observe three large peaks far from the mean which drive the disassortativity. These are due to hubs connecting to a large number of low degree nodes. The network is shown in fig. 2.8 where the yellow edges and their associated vertices are in the range $133 \leq |\Delta| \leq 122$ and the green edges and their vertices have $66 \leq |\Delta| \leq 62$. This allows us to easily identify the vertices and edges that strongly contribute to the disassortativity.

For the mythological networks presented in this work, some narratives have similar assortativity values but different mechanisms can account for this. The $\psi(\Delta)$ distribution allows us to visualise the assortativity to show the differences in interactions.

### 2.8.2 Further Similarity Measures

Degree assortativity measures the similarity of the degrees of vertices connected to one another. Assortativity can also be used for other properties. One may define the *clustering assortativity* $r_C$ by replacing the degree $k_i$ of vertex $i$ in eq. (2.65) with its clustering coefficient $C_i$ from eq. (2.39).

Further measures of similarity have also been defined for networks (Newman 2009). The *Pearson similarity* measure is also used here. This is similar to Pearson correlation coefficient of eq. (2.65) but instead of examining the degrees of vertices that are linked, it measures how many neighbours $v$ each pair of vertices $i, j$ share. Here $i$ does not have to share an edge with $j$. This type of similarity measure is often referred to as structural equivalence.

The Pearson similarity for vertices $i$ and $j$ is defined by

$$r_{ij} = \frac{\sum_v (A_{iv} - N^{-1}k_i)(A_{jv} - N^{-1}k_j)}{\sqrt{\sum_v (A_{iv} - N^{-1}k_i)^2}\sqrt{\sum_v (A_{jv} - N^{-1}k_j)^2}}. \tag{2.68}$$

The average Pearson similarity $r_P$ for the network can be calculated by summing over all $r_{ij}$ for all pairs of vertices.

A network that is highly disassortative may have positive Pearson similarity as many low degree vertices that do not share an edge may be connected to the same high degree vertices. For example a star graph is a graph where $N-1$ vertices have degree $k = 1$ and one vertex has degree $k = N - 1$. For $N = 100$ for example, this has a degree assortativity of $r = -1$ but has a positive Pearson similarity of $r_P = 0.96$.

## 2.9 Centrality Measures and Robustness

In a network, it is often useful to identify influential or central nodes. One simple measure of centrality is the degree. A vertex with a high degree is likely to have more influence on the properties of the network than a vertex with a low degree.

Another measure of influence is the *betweenness centrality* of a vertex. This is the total number of geodesics that pass through a given vertex (Freeman 1977). A vertex with a high betweenness centrality has a high probability to be on a shortest path between two other vertices. Therefore it controls the flow of information between other vertices.

If $\sigma(i, j)$ is the number of geodesics between vertices $i$ and $j$, and if the number of these which pass through node $l$ is $\sigma_l(i, j)$, the betweenness centrality of vertex $l$ is

$$g_l = \frac{2}{(N-1)(N-2)} \sum_{i \neq j} \frac{\sigma_l(i, j)}{\sigma(i, j)}. \tag{2.69}$$

The normalization ensures that $g_l = 1$ if all geodesics pass through $l$.

An expression analogous to Eq. (2.69) can be developed for edges to determine the *edge betweenness centrality* which we will return to in the next section. This determines which edges are traversed frequently on shortest paths in a network.

To test the *robustness* of a network, vertices can be removed and the size of

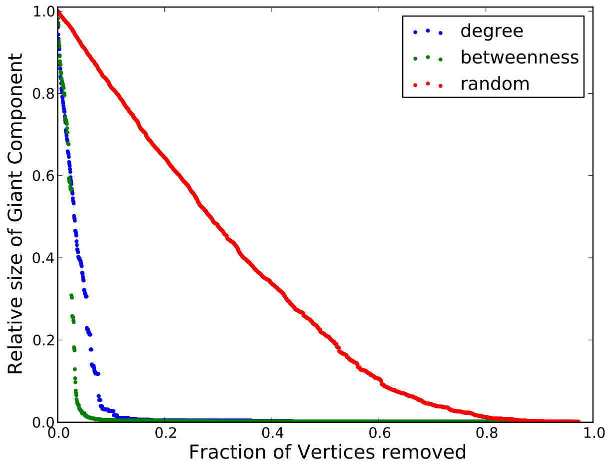



*Fig. 2.9:* The robustness for the network of the protein interaction network in yeast. The network is very susceptible to targeted attacks on degree (blue) and betweenness (green) but is robust to the random removal of vertices (red).

the giant component measured. If the giant component fragments quickly then the network is not robust (Albert *et al.* 2000). Targeted removal of vertices involves eliminating vertices with the highest degree or betweenness centralities. When removing vertices by betweenness centrality, the betweenness must be recalculated after each removal. Vertices can also be removed by selecting them randomly. Networks that have scale-free degree distributions tend to be robust to random removal of vertices but fragile to targeted removal of vertices (Albert *et al.* 2000).

Fig. 2.9 shows the robustness for the protein interaction network in yeast (Jeong *et al.* 2001). The giant component breaks down quickest by removing vertices of high betweenness centrality followed closely by the removal of vertices by degree. Vertices were removed randomly 50 times and the average size of the giant component are shown as the red points. This network is robust when vertices are removed randomly.

In the context of the social networks presented in this work, a lack of robustness indicates the network is overly reliant on a few characters. Naïvely, we would expect a social network to be robust due to a lack of disassortativity. For narratives however, this may not be the case as they are often centred on a single protagonist. Testing the robustness provides information as to whether the tale is hero-centred or society-centred.

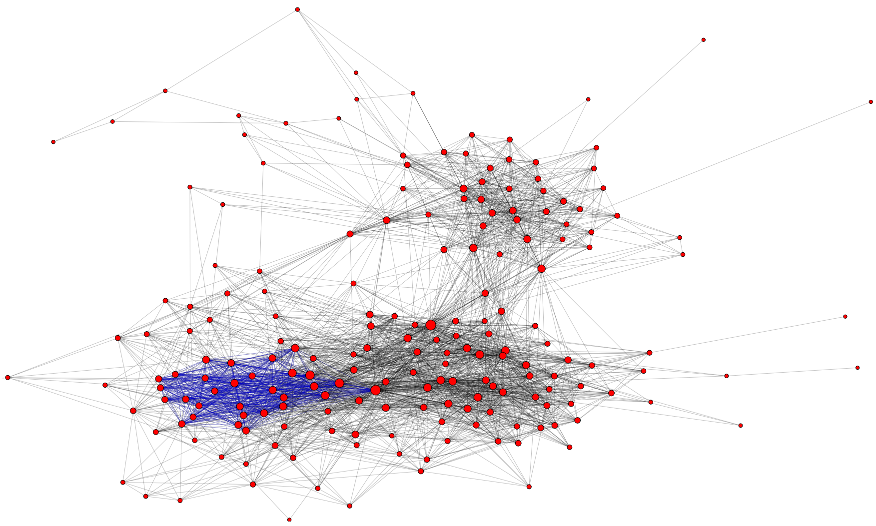


*Fig. 2.10:* A network of jazz musicians between 1912 and 1940 (Gleiser & Danon 2003). There are two distinct communities in this network due to the racial segregation present at the time. Though collaborations did exist, bands mostly comprised exclusively black or white musicians. The blue coloured edges here represent a clique of 31 musicians who all collaborated together at one time or another.

## 2.10 Communities

A *clique* is maximal complete subgraph within a network containing at least three vertices (Luce & Perry 1949). As complex networks tend to have high clustering coefficients (Watts & Strogatz 1998), a large number of cliques of size $C_s = 3$ (triangles) are contained within them. Larger cliques are found in social networks though their number decays rapidly.

The social network of Jazz musicians (Gleiser & Danon 2003) contains one particularly large clique containing 31 vertices. Fig. 2.10 shows this as the cluster with blue edges. Each musician collaborated with every member of that clique at one time or another. However, the network contains 738 cliques in total only two of which comprise 20 or more members. Fig. 2.11 depicts the complementary cumulative the clique size distribution $P_{C_s}$ with a fitted Gaussian distribution. With the exception of the two largest cliques, the distribution fits the data well signalling the speed of decay.

The requirement for every member of a clique to be connected is quite a stringent one. Instead more focus is put on a *community* within a network rather than a clique. A community in a network is a densely connected group of vertices with

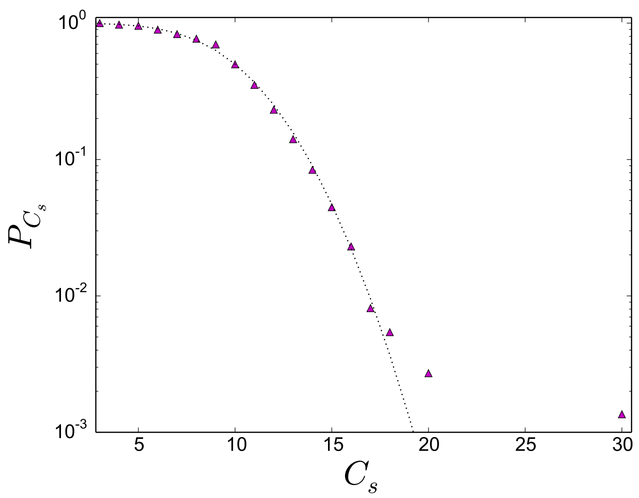



*Fig. 2.11:* The clique size distribution for the network of Jazz musicians (Gleiser & Danon 2003). The dotted black line represents a Gaussian distribution. The rightmost two points which are not well fitted by the distribution represent only two cliques of the 738 cliques.

relatively few edges going from vertices within that cluster going to vertices outside of that cluster (Newman & Girvan 2004). However, there is no universally accepted definition of a community; this often depends on the system or application at hand (Fortunato 2010).

Fig. 2.10 shows a network of jazz musician who collaborated together between 1912 and 1940 (Gleiser & Danon 2003). Due to racial segregations present in this era, bands almost exclusively contained black or white musicians (though collaborations did exist). Gleiser & Danon (2003) identified two major communities (as shown in the upper right cluster and the larger lower cluster in fig. 2.10). They also identified two communities within the larger cluster, distinguishing between musicians based in Chicago and musicians based in New York.

### *2.10.1 Community Detection*

In order to identify communities within a graph, the Girvan-Newman algorithm is employed (Girvan & Newman 2002). This algorithm removes edges with the highest betweenness as these tend to be the most "between" communities. After each edge removal, the edge betweenness needs to be recalculated. This breaks the network down into smaller sub-components as it progresses. To find the optimal number of communities the *modularity* of these sub-components within the

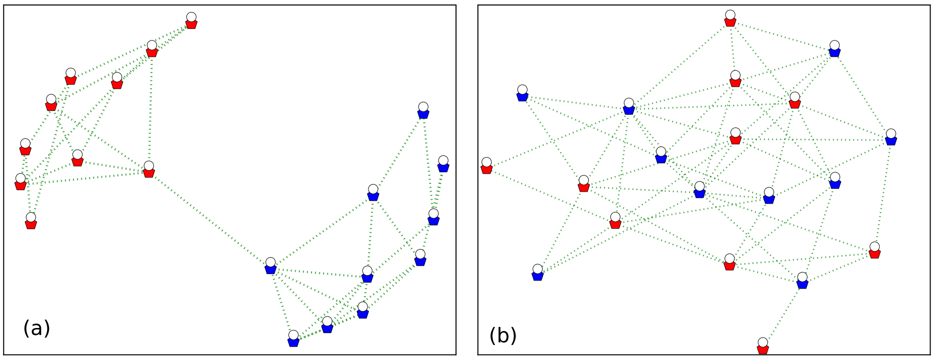



*Fig. 2.12:* Illustration of communities in networks with different modularity. Panel (a) has two types of nodes and two visible communities. The modularity for two communities $Q = 0.47$. Panel (b) again has two types of nodes but they are mixed randomly. This has modularity $Q = -0.12$ for two communities.

entire network is evaluated (Newman & Girvan 2004). The modularity reaches a maximum (or sometimes a plateau) at the optimal number of communities.

To determine the modularity, the $n \times n$ matrix **e** is constructed, where $n$ is the number of communities. The element $e_{st}$ of the matrix is the fraction of all edges in the network that link nodes in community $s$ to nodes in community $t$. Denoting $a_t$ as the sum of each row, $a_t = \sum_s e_{st}$. The modularity $Q$ is then defined by

$$Q = \sum_t (e_{st} - a_t^2). \tag{2.70}$$

If the number of edges between communities is no better than if they were randomly distributed, then $Q = 0$. At the other extreme, if the network is partitioned into $n$ sparsely inter-connected communities each containing approximately $L/n$ edges, then $E_{st} \approx \delta_{st}/n$ and $a_t \approx E_{tt} \approx 1/n$, so that $Q \approx 1 - 1/n$. Although modularity is bounded by $Q = 1$ for large $n$, it is typically between about $0.3$ and $0.7$ in social networks with varying degrees of community structure (Newman & Girvan, 2004).

Fig. 2.12 gives an example of two networks each with two types of nodes. In fig. 2.12 (a) the two communities almost exclusively mix with members of their own community. This yields a modularity of $Q = 0.47$. In fig. 2.12 (b) they mix randomly and a value of $Q = -0.12$ is obtained.

# 3. SOCIAL NETWORKS

A *social network* is a graph in which the vertices represent people and edges represent some type of interaction between them. Newman & Park (2003) demonstrate that social networks are different to other types of complex networks. They outline two key differences. The first is that social networks are usually assortatively mixed by degree, whereas non-social networks are almost exclusively disassortative. The second is that the transitivity $C_T$ is higher than one would expect given the degree distribution. A naïve estimate for the transitivity $C_\mathrm{n}$ is obtained from eq. (2.54). For social networks, $C_T > C_\mathrm{n}$, Newman & Park (2003) find that this tends not to be the case for non-social networks.

Newman & Park (2003) explain both of these observations by fact that social networks are usually divided into communities. Šubelj & Bajec (2012) find that real world social networks are also assortatively mixed by clustering coefficient and that this also indicates the presence of communities. Note that community structure is not exclusive to social networks; for example biological, technological and economic networks also exhibit community structure (Fortunato 2010).

Social networks are also found to be small world (Amaral *et al.* 2000). This property is often used in epidemiology to model disease transmission on networks (Kuperman & Abramson 2001).

Therefore we expect that social networks exhibit the following properties:

- Small worldness; $\ell \approx \ell_\text{rand}$, $C \gg C_\text{rand}$
- Assortativity; $r_k > 0$, $r_C > 0$
- High clustering coefficient; $C_T > C_\mathrm{n}$
- Community structure

Non-social networks may exhibit some of these properties but the combination of all seems exclusive to social networks in empirical studies carried out so far.

## 3.1 Types of Social Networks

Examples of social networks include the networks of school students (Resnick *et al.* 1997), movie actors (Barabási & Albert 1999), scientific co-authors (Newman 2001b), company boards of directors (Davis *et al.* 2003), e-mail contacts (Guimera *et al.* 2003), romantic relationships between highschool students (Bearman *et al.* 2004), characters in Victorian novels (Elson *et al.* 2010), mobile phone users (Palchykov *et al.* 2013), as well as a whole range of online social networks (Mislove *et al.* 2007) to name but a few. There is also a field of study dedicated to non-human social networks (Croft *et al.* 2008) ranging from the interactions of baboons (Dunbar & Dunbar 1974) to wolves (Cohen *et al.* 2013).

Due to this diverse range of datasets, some authors claim that not all of these networks are social networks. For example, Amaral *et al.* (2000) argue that the movie actor network or scientific collaboration networks are 'economic networks' rather than social networks. Amaral *et al.* (2000) go on to show that degree distribution of the social networks of Utah mormons (Bernard *et al.* 1988) and school friendships (Fararo & Sunshine 1964) follow Gaussian distributions whereas the movie actor network follows a power law (Barabási & Albert 1999). They attribute different degree distribution behaviours to different classes of small-world networks. However, the network of mormons only contains 43 vertices which is a small sample size making it difficult to determine the nature of the degree distribution.

Newman (2001a), on the other hand, claims that collaboration networks are truer examples of social networks than the friendships of school students. He argues that in a school, children may give different answers as to who their friends on different days. However, in a scientific collaboration network, if two authors collaborate they are almost certainly acquainted (with the exception of subjects such as experimental physics where author lists can be vast).

Similarly, sexual contact networks have very different properties to other social networks. Bearman *et al.* (2004) observe that the sexual network of a group of high-school students is disassortative. This network has 573 vertices but it only contains one triangle giving it a very low clustering coefficient. By the properties listed above this does not behave like the Newman & Park (2003) definition of a social network. Newman (2003b) and Newman & Park (2003) do, however, classify it as a social network. They also show that the error in the assortativity from eq. (2.66) is higher than its value indicating it is not truly disassortative.

Therefore different types of social networks can have differing properties depending on how edges are created. To investigate this further, six different types of social networks are analysed here. Two examples were chosen for each, based on availability of data to represent each category. These are (i) the *physical co-location* networks of students at Faux Mesa high school (Resnick *et al.* 1997; Hunter *et al.* 2008) and the face-to-face contact of participants at an infectious diseases exhibition in Dublin's Science Gallery (Isella *et al.* 2011); (ii) the *collaboration* networks of authors on the condensed matter *arXiv* from 1995 to 1999 (Newman 2001b) and American jazz musicians from 1912 to 1940 (Gleiser & Danon 2003); (iii) the *communication* networks of emails sent at Enron (Klimt & Yang 2004; Leskovec *et al.* 2009) and the PGP (Pretty Good Privacy) web of trust (Boguñá *et al.* 2004); (iv) the *online social networks* of a sample of Facebook users (Viswanath *et al.* 2009) and users of the technology news site slashdot (Kunegis *et al.* 2009; KONECT 2013); (v) the *sexual contact* networks of Jefferson high school students (Bearman *et al.* 2004) and a Brazilian sexual escort network (Rocha *et al.* 2011); and (vi) the *fictional* networks based on the co-appearance of characters in Victor Hugo's *Les Misérables* (Knuth *et al.* 1993) and Marvel Universe's comic book characters (Alberich *et al.* 2002). The core properties of these networks are listed in table 3.1.

### *3.1.1 Degree Distributions*

The degree distributions for the 12 networks are shown in fig. 3.1. Of the three classes of complex networks Amaral *et al.* (2000) identifies (power law, truncated power law, and no power-law regime, see section 2.2.2), all belong to the third category; namely distributions with fast decaying tails and no power law regime. However, within this we find four different distributions:

- the physical co-location network of Faux Mesa high school, the sexual contact network of Jefferson high school students and the fictional network of *Les Misérables* are all best fitted by exponential distributions.

- the collaboration network of jazz musicians in fig. 3.1 (c) is shown with a fitted Gaussian distribution from eq. (2.23). However the $AIC_c$ and BIC give almost equal weights to a Weibull distribution from eq. (2.26) suggesting either model is a good candidate.

- the physical co-location network of participants at an infectious diseases ex-

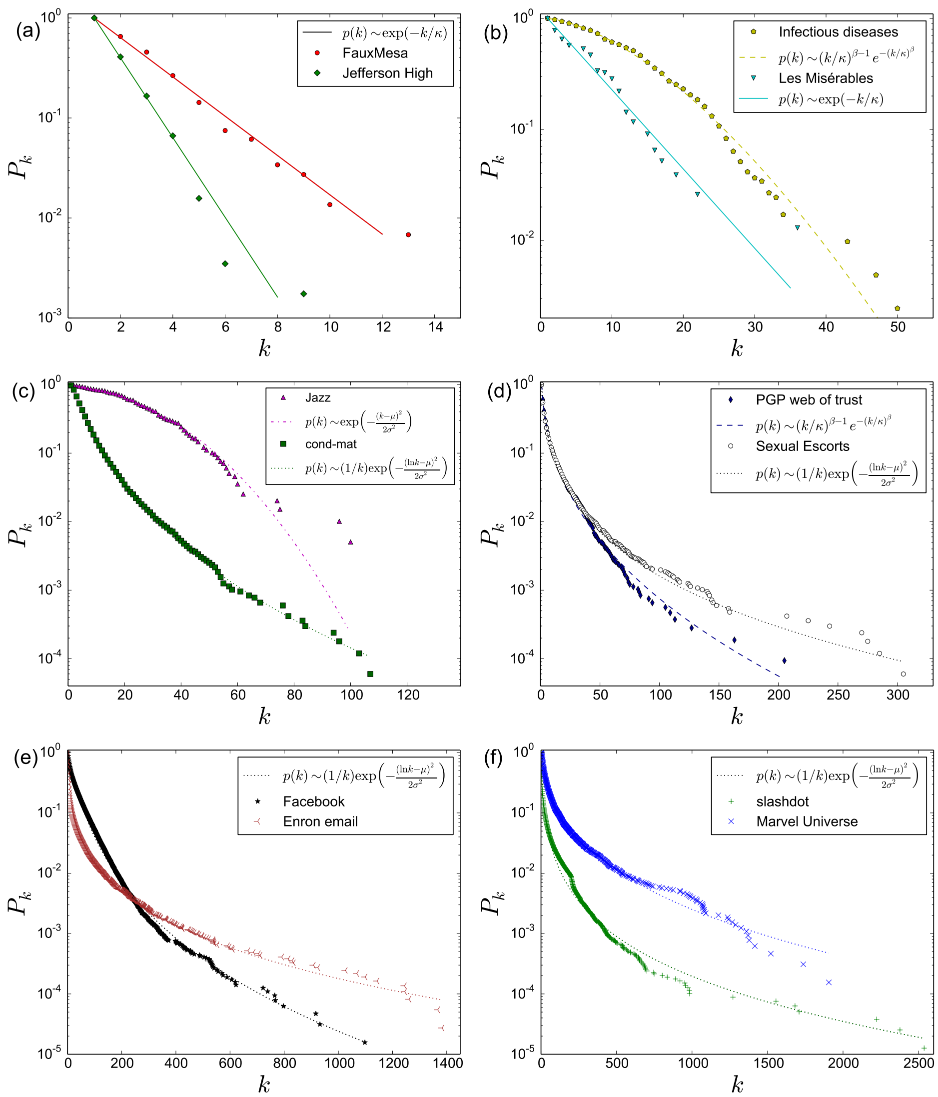



*Fig. 3.1:* The complementary cumulative degree distributions for the 12 social networks shown on a semi-log scale. Four different kinds of distributions are shown here; continuous lines represent exponential distributions, dashed lines represent Weibull distributions, dotted lines represent log-normal distributions and the dashed-dotted line in panel (c) represents a Gaussian distribution.

hibition and the communication network of the PGP web of trust are both best fitted by Weibull distributions. For the infectious diseases network, the parameter $\beta > 1$ causing the distribution to be concave down after the maximum. This is not the case for the PGP web of trust.

- the collaboration network of co-authors on the cond-mat arXiv, the communication network of Enron employees' emails, the online social networks of Facebook and slashdot users, the network of Brazilian sexual escorts and the fictional network of Marvel Universe characters are all well fitted by log-normal distributions from eq. (2.24). However, the $AIC_c$ and BIC also gave high weights to a Weibull distribution for Facebook and a truncated power law for slashdot.

The debate about the applicability of power laws in complexity science extends beyond the study of networks. Clauset *et al.* (2009) show that a large number of empirical data sets that have been fitted by power-law distributions are better fitted by log-normal distributions. This claim of log-normals being better suited to empirical data than power-laws has also been made for urban growth (Eeckhout 2004) and firm growth in economics and econophysics (Mansfield 1962). For example Eeckhout (2004) shows that despite Zipf's law being applied to the growth of cities for more than half a century, it only holds for the upper tail of the distribution. The exponent of the distribution is sensitive to where the truncation is chosen. He shows that a log-normal distribution is a better candidate model and the parameters are far more robust.

A further similarity between log-normal distributions and power laws is that in each case the growth rate can be independent of the system size (Fujiwara *et al.* 2004). In complex networks, the degree is, of course, bounded by the system size, however the maximum degree is often found to be orders of magnitude higher than the average degree (see table 3.1 for example). For further details on the debate about the suitability and ubiquity of power laws, see (for example) Edwards *et al.* (2007); Clauset *et al.* (2009); Stumpf & Porter (2012).

### *3.1.2 Further Network Properties*

(i) *Physical co-location networks*

Physical co-location networks both have relatively low average and maximum degrees. The variance is somewhat large for the infectious diseases network but its standard deviation is still less than its mean. They are both degree and clustering assortative. They also have larger clustering coefficients than the their random and naïve counterparts. However in both cases $\ell > \ell_{\text{rand}}$. Using the small-worldness test of Humphries & Gurney (2008) from eq. (2.57), we find $S > 1$. However, strictly speaking, they do not meet the Watts-Strogatz definition of small world.

*Tab. 3.1:* Comparison of network properties for 6 types of social networks. The number of vertices is given by $N$, $L$ is the number of edges, $\langle k \rangle$ is the mean degree, $k_{\text{max}}$ is the largest degree in the network, $\langle k^2 \rangle$ is the mean of the square of the degree (related to the variance), $r_k$ is the degree assortativity, and $r_C$ is the clustering assortativity. In the lower half of the table $\ell$ is the average path length, $C$ is the clustering coefficient ($\ell_{\text{rand}}$ and $C_{\text{rand}}$ are their random equivalents). The diameter, or longest shortest path is $\ell_{\text{max}}$, $C_T$ is the transitivity and $C_{\text{n}}$ is the naïve expectation for the transitivity. The symbol * signals that the error calculated from eq. (2.66) is larger than this value indicating that there are no degree-degree correlations.

| | $N$ | $L$ | $\langle k \rangle$ | $k_{\text{max}}$ | $\langle k^2 \rangle$ | $r_k$ | $r_C$ |
|---|---|---|---|---|---|---|---|
| | Physical co-location | | | | | | |
| Faux Mesa high school | 147 | 202 | 2.75 | 13 | 11.70 | 0.12 | 0.36 |
| Infectious diseases | 410 | 2765 | 13.49 | 50 | 252.43 | 0.23 | 0.34 |
| | Collaboration | | | | | | |
| cond-mat arXiv | 16726 | 47594 | 5.69 | 107 | 73.57 | 0.19 | 0.20 |
| Jazz Musicians | 198 | 2742 | 27.70 | 100 | 1070.24 | 0.02* | 0.01* |
| | Communication | | | | | | |
| Enron email | 36692 | 183831 | 10.02 | 1383 | 1403.62 | -0.11 | 0.19 |
| PGP web of trust | 10680 | 24316 | 4.55 | 205 | 85.98 | 0.24 | 0.50 |
| | Online Social Network | | | | | | |
| Facebook | 63731 | 817090 | 25.64 | 1098 | 2257.25 | 0.18 | 0.24 |
| slashdot | 79120 | 467869 | 11.83 | 2537 | 1731.86 | -0.07 | 0.26 |
| | Sexual Contact | | | | | | |
| Jefferson high school | 573 | 477 | 1.67 | 9 | 3.78 | -0.03* | 0.49 |
| Sexual Escorts | 16730 | 39044 | 4.67 | 305 | 130.98 | -0.11 | 0.00* |
| | Fiction | | | | | | |
| Les Misérables | 77 | 254 | 6.60 | 36 | 79.53 | -0.17 | 0.27 |
| Marvel Universe | 6445 | 168267 | 52.22 | 1906 | 15857.53 | -0.16 | - 0.11 |

| | $\ell$ | $\ell_{\text{rand}}$ | $\ell_{\text{max}}$ | $C$ | $C_{\text{rand}}$ | $C_T$ | $C_{\text{n}}$ |
|---|---|---|---|---|---|---|---|
| | Physical co-location | | | | | | |
| Faux Mesa high school | 6.81 | 4.87 | 16 | 0.25 | 0.02 | 0.28 | 0.03 |
| Infectious diseases | 3.63 | 2.59 | 9 | 0.46 | 0.03 | 0.44 | 0.06 |
| | Collaboration | | | | | | |
| cond-mat arXiv | 6.63 | 5.76 | 18 | 0.62 | 0.00 | 0.36 | 0.00 |
| Jazz Musicians | 2.24 | 1.92 | 6 | 0.62 | 0.14 | 0.52 | 0.26 |
| | Communication | | | | | | |
| Enron email | 4.03 | 4.81 | 13 | 0.50 | 0.00 | 0.09 | 0.06 |
| PGP web of trust | 7.49 | 6.24 | 24 | 0.27 | 0.00 | 0.38 | 0.01 |
| | Online Social Network | | | | | | |
| Facebook | 4.32 | 3.73 | 15 | 0.22 | 0.00 | 0.15 | 0.01 |
| slashdot | 4.04 | 4.83 | 12 | 0.06 | 0.00 | 0.02 | 0.02 |
| | Sexual Contact | | | | | | |
| Jefferson high school | 15.89 | 11.83 | 37 | 0.00 | 0.00 | 0.01 | 0.00 |
| Sexual Escorts | 5.79 | 6.44 | 17 | 0.00 | 0.00 | 0.00 | 0.01 |
| | Fiction | | | | | | |
| Les Misérables | 2.64 | 2.50 | 5 | 0.57 | 0.09 | 0.50 | 0.24 |
| Marvel Universe | 2.64 | 2.57 | 6 | 0.77 | 0.01 | 0.20 | 0.27 |

They do however fulfil the Newman & Park (2003) criteria for social networks.

(ii) *Collaboration networks*
The two collaboration networks have quite different properties. The network of co-authors on the condensed matter arXiv has a relatively low average degree, maximum degree and variance. It is assortative, as is commonly found in collaboration networks (Newman 2002). It also has a very high clustering coefficient and transitivity compared to its random and naïve clustering coefficients.

In the network of Jazz musicians on the other hand, vertices are linked to an average of one eighth of the other musicians in the network. The musician with the highest degree performed with half the musicians active during that period and the variance in degree is large. Its assortativity is $r_k = 0.02 \pm 0.02$ indicating there no significant degree-degree correlations. The clustering coefficient and transitivity are still high, though not significantly larger than their random and naïve versions.

Despite both these being collaboration networks however, edges are not formed in the same way. For the network of jazz musicians, if two bands record together, everybody in each band is linked. Similarly who performs with whom is not entirely up to each individual musician. At the time there was also quite a small pool of musicians available to perform with. In scientific collaboration however, two co-authors actively choose (in general) to work together. There is also less external pressure to collaborate with a specific scientist. In this regard, there is more freedom for a vertex to choose their neighbours than in the case of jazz musicians. There is also a much larger choice of vertices to interact with.

(iii) *Communication networks*
The maximum degree in the Enron email network is very large at $k_{\max} = 1383$. There is also a large variance in the degree. In fig. 3.2 (a), the $\psi(\Delta)$ distribution from section 2.8.1 for the difference in degrees of vertices at either end of an edge is shown. At each end of the plot, we see there is a high fraction of edges with $|\Delta| = 1382$. Hence the vertex with the highest degree interacts with multiple vertices which interact with no other nodes in the network. As a result the network is disassortative. This network is small world, however the transitivity is comparable to its naïve estimate. This network does not meet the Newman & Park (2003) definition of a social network.

The PGP web of trust network on the other hand has a relatively low maximum degree and a low variance in degree. It is assortative and we observe in fig. 3.2 (b)

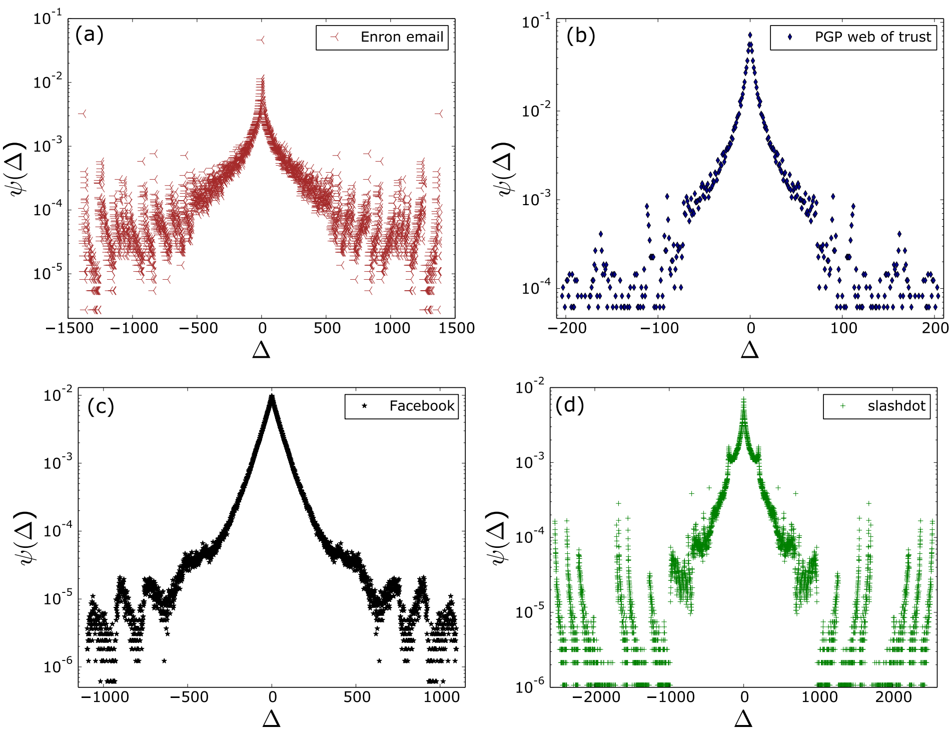



*Fig. 3.2:* The $\psi(\Delta)$ distribution showing the difference in degrees of vertices at either extremity of an edge. Panel (a) shows the Enron email network. This network is disassortative and there are numerous edges with a difference of degrees of $|\Delta| = 1382$. In contrast the PGP web of trust $\psi(\Delta)$ distribution is shown in panel (b). This is assortative and has a much more uniform distribution. Panel (c) displays the Facebook users network which is also assortative. Panel (d) however is for the online social network of slashdot users. This network is disassortative and contains a large number of fluctuations as $|\Delta|$ increases.

that there are comparatively few fluctuations in the difference of degrees of vertices at the ends of an edge. Like the physical co-location networks and the network of co-authors on the condensed matter arXiv, it has $\ell > \ell_{\text{rand}}$, $C \gg C_{\text{rand}}$ and $C_T \gg C_{\text{n}}$.

It is perhaps misleading to group the two communications networks together as their edges are created by very different mechanisms. The PGP web of trust requires both users to trust each other with their encryption keys in order for an edge to be formed. As a result this is more like a collaboration network. In the Enron email network however, an edge is created if one individual emails another. There is no trust element involved and there is no requirement for the communication to go both ways. Technically this is a directed network but we treat each network as undirected in this analysis as edges in social networks are preferred to be mutual. These factors could explain why the two networks yield such different properties.

(iv) *Online Social networks*

The two online social networks are the largest networks used in this analysis. They too have low average degrees relative to their size. Facebook is assortative while slashdot is not. The $\psi(\Delta)$ distribution of each is shown in fig. 3.2 (c-d). There are large fluctuations around $|\Delta| \approx 1500$ and $|\Delta| \approx 2500$ for the slashdot users. However the distribution for Facebook users is much more uniform. Facebook also has a larger transitivity than the naïve expectation from its degree properties unlike that of the slashdot network. They are however both small world and clustering assortative.

Once again, these two networks have different mechanisms for creating edges. On Facebook users are only friends if both parties agree. It only takes one user to initiate but the other must accept for a link to be formed. This is not the case for slashdot. Here users can mark any other user as a friend or foe. There is no requirement for the other user to reciprocate. Hence these are two different types of online social networks.

Probing this further, we look at other studies of online social networks. A different Facebook network of 4,039 users (McAuley & Leskovec 2012) is also assortative ($r_k = 0.06$). A network of 23,628 Google+ users (McAuley & Leskovec 2012) however is disassortative ($r_k = -0.23$), as is a network of 81,306 Twitter users ($r_k = -0.04$) (McAuley & Leskovec 2012). Neither Google+ nor Twitter require reciprocation for edges to be created. Any Google+ user can add another user to their circles without being added in return. Similarly any Twitter user can follow another Twitter user without the other following them back. Again these networks are technically directed social networks rather than undirected ones. In some cases only undirected data are available.

Kumar *et al.* (2010) find that online social networks go through distinct periods of growth characterised by changes in the diameter and the structure of the giant component. This may perhaps account for the different properties observed in them as each of these websites were created at different times. However, it seems more likely that the differences observed here are due to the nature of how the edges are created.

(v) *Sexual Contact networks*

The Jefferson high school sexual network has a relatively low average degree, maximum degree and variance. Its assortativity is $r_k = -0.03 \pm 0.04$ indicating there are no significant degree-degree correlations. This network is not small world and contains one triangle due to two bisexual vertices.

The network of Brazilian sexual escorts has a large maximum degree. This network is disassortative. Though its average path length is shorter than its random counterpart, it contains no triangles. Hence it is technically not small world.

These two networks are (for the most part) bipartite graphs as there are two distinct types of vertices that do not interact with one another. For the case of Jefferson high school there is roughly an equal number of male and female nodes. However in the escort network there are far more users than escorts. This likely leads to a larger variance in degree and the disassortativity. Though these networks are technically social networks, they too have distinct mechanisms for creating edges and different types of vertices. As a result, it is unsurprising that they have such different properties to the other types of network analysed here.

(vi) *Fictional networks*
Both fictional networks have similar average and maximum degrees relative to their size. They are also both disassortative by degree. The network of characters in *Les Misérables* however is clustering assortative unlike that of the Marvel Universe. Both these networks are small world, however in each case the their transitivity is close to that of the naïve estimate. Therefore both of these networks do not meet the criteria of Newman & Park (2003) for social networks.

*Friendly and Hostile networks*
Finally we examine and compare the properties of a *friendly network* and a *hostile network*. The data comes from the network of slashdot users where, as mentioned above, users can mark other users as friends or foes. The core properties of each network can be found in table 3.2.

Users mark other users as friends almost twice as frequently as foes. The full slashdot network has a degree assortativity of $r_k = -0.07$. Here we observe that the friendly network has a slightly higher assortativity of $r_k = -0.06$, but the hostile network has a considerably lower assortativity of $r_k = -0.16$. Neither network has a large clustering coefficient. In particular however, the network of hostile users has a lower transitivity than naïvely predicted. This lack of triangles is related to the notion of structural balance. This was also observed in the hostile network of

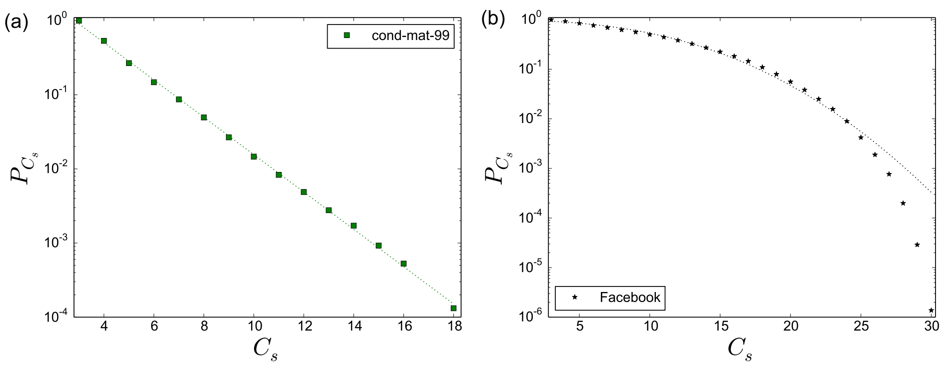



*Fig. 3.3:* The clique size distribution $P_{C_s}$ for the two social networks. Panel (a) shows the distribution for the cond-mat arXiv with a fitted exponential. Panel (b) is the clique sizes distribution for the network of Facebook users with a fitted Gaussian distribution.

Szell & Thurner (2010). There, the online social network of a multiplayer online game is analysed which is also disassortative and contains few triangles.

### 3.1.3 Community Structure

With the exception of the sexual contact networks, these social networks contain a number of different sized cliques. Fig. 3.3 shows the fraction of clique sizes $P_{C_s}$ plotted against the size of a clique $C_s$ for two of the social networks. In the case of co-authors on the condensed-matter arXiv this distribution is exponential, as shown in fig. 3.3 (a). This however is not the case for most of the others distributions, for example fig. 3.3 (b) shows the clique size distribution for the Facebeook network with a fitted Gaussian distribution.

*Tab. 3.2:* Comparison of the friendly network and hostile network of slashdot where users can mark other users as friends or foes. The hostile network is more disassortative and contains almost no triangles.

| | $N$ | $L$ | $\langle k \rangle$ | $k_{\text{max}}$ | $\langle k^2 \rangle$ | $r_k$ | $r_C$ |
|---|---|---|---|---|---|---|---|
| slashdot friendly | 69998 | 351013 | 10.03 | 2506 | 1290.04 | -0.06 | 0.34 |
| slashdot hostile | 37412 | 118755 | 6.35 | 685 | 412.67 | -0.16 | 0.06 |

| | $\ell$ | $\ell_{\text{rand}}$ | $\ell_{\text{max}}$ | $C$ | $C_{\text{rand}}$ | $C_T$ | $C_{\text{n}}$ |
|---|---|---|---|---|---|---|---|
| slashdot friendly | 4.16 | 5.09 | 13 | 0.06 | 0.00 | 0.03 | 0.02 |
| slashdot hostile | 4.23 | 5.89 | 13 | 0.01 | 0.00 | 0.01 | 0.02 |

In the case of the sexual networks we use the Girvan-Newman algorithm (Girvan & Newman 2002). For both of these networks the algorithm finds communities with a high modularity. Therefore all these social networks exhibit some form of community structure.

### *3.1.4 Summary*

The 12 social networks analysed here share some common properties. They each exhibit community structure and they all fall into the third class of small-world network that Amaral *et al.* (2000) describe with fast decaying tails and an absence of a power-law regime (though technically some are not small world). Log-normal distributions are found to be the most common distribution. Every network with a maximum degree above 250 is well fitted by a log-normal distribution. However, in other properties they differ depending on how their edges are formed.

In the case where edges are mutual and freely chosen, specifically for the two physical co-location networks, the collaboration network of arXiv co-authors, the communication network of PGP web of trust users and the online social network of Facebook, the networks have similar properties. These all are assortative, have a high clustering coefficient and transitivity and have an average path length $\ell \gtrsim \ell_{\mathrm{rand}}$. These meet the Newman & Park (2003) classification of social networks.

The following networks have different properties. In the collaboration network of jazz musicians there is less freedom for an individual to choose who they are linked to. This network then has a high average degree and almost no assortativity. The networks in which the edges do not have to be mutual, such as in the case of emails or slashdot users, are disassortative with transitivities not deviating from their naïve values. Sexual contact networks have almost an absence of triangles and are not assortative. These networks also contain two types of vertices. Finally networks appearing in the fictional works analysed here are disassortative and have comparable clustering coefficients to their naïve versions. Networks of hostile interactions have different properties to their friendly counterparts.

Based on the properties of the networks with reciprocating edges, we expect social networks to have the following properties:

- Fast decaying degree distribution with no power-law regime
- Close to small world; $\ell \gtrsim \ell_{\mathrm{rand}}$, $C \gg C_{\mathrm{rand}}$

- Assortativity; $r_k > 0$, $r_C > 0$
- High clustering coefficient (transitivity); $C_T > C_{\mathrm{n}}$
- Community structure.

# 4. MYTHOLOGY

The word “myth” is derived from the Greek word “mythos” meaning “story”. In particular, it is a story concerning the distant past. Mythology is a collection of myths, and though there is no single definition of a myth, they are most commonly described as religious or sacred narratives (Eliade 1998; Campbell 2001; Dundes 1984; Leeming 2005; Schrempp & Hansen 2002).

Many authors make specific distinctions between different types of myths. Nilsson (1932), for example, distinguishes between divine and heroic mythology. Divine mythology concerns the gods and what he calls cult myths. Heroic myths however often begin with folktales and end with incidents that may have a historical appearance.

Bascom (1965) identifies three different types of prose narrative; folktales, myth and legend. He defines them as follows:

- folktales are narratives that are regarded as fiction. They are often “timeless” and generally not to be taken seriously.
- Myths may be considered by the society to be truthful accounts of what happened in the remote past. They are usually sacred.
- Legends are set in a period less remote than myths but are also considered basically true by the narrator and audience. They tend to be more secular than sacred.

However, even the lines between these distinctions are often blurred. Dundes (1997) in his discussion on the debate between Claude Lévi-Strauss and Vladimir Propp on the approach to structuralism criticises the former for classifying Oedipus as a myth. He calls it a folktale, citing the Bascom (1965) definition and cites its Aarne-Thompson classification number. However, by the Nilsson (1932) definition it is a heroic myth. Leeming (1991) similarly calls it a heroic myth. Using the Bascom (1965) definition, some elements of Oedipus could allow one to interpret it as a legend also. For example unlike most folktales it is not timeless, it has a distinct setting in both place and time. Some elements of it are also believed to be historical, for example the plague described in it is likely based on a historical epidemic (Kousoulis *et al.* 2012).

Instead of distinguishing between types of prose, Leeming (1991) identifies four types of myth. These are:

- cosmic myths (i.e. dealing with creation)
- myths of the gods (such as the Greek Pantheon)
- hero myths (similar to the definition of Nilsson (1932) above)
- place and object myths (e.g. myths about Atlantis or the Golden Fleece).

Campbell (2001) and Leeming (2005) argue that religion can be seen as a system that lends authority to myths. However, myths often serve many other purposes outside of religion. Some myths or legends are associated with place-naming or topography, this is a particularly common feature in the Irish *Táin Bó Cuailnge* (Kinsella 1969, p. xiii). Although the Táin is not a "sacred narrative", it is regarded as myth (O'Rahilly 1946; Kinsella 1969; Leeming 2005).

With all these differing types of narratives and opaque distinctions between them, we use the word "myth" in the broadest possible sense. The Oxford dictionary definition of *myth* is "a traditional story, especially one concerning the early history of a people or explaining a natural or social fact." This definition is close to the original meaning of the word and this is the context to which it will be used here. Using methods of network theory, we seek to uncover differences between myths quantitatively rather than qualitatively, like those of Nilsson (1932) or Leeming (1991) above.

## 4.1 Comparative Mythology

There are a large number of approaches to the study of mythology. A common approach involves searching for meaning in myths to explain phenomena, such as everyday occurrences like the movement of the sun across the sky (Leeming 1991) to more irregular events like a natural disasters, e.g. a flood. A brief overview of a selection of types of comparative mythology is presented here. A more detailed overview of comparative mythology can be found in the introduction of Morford & Lenardon (1999). Dundes (1984) provides a collection of articles on differing approaches to the analysis of myth and Schrempp & Hansen (2002) offer an assessment of the present state of the study of myth.

An early approach to comparative mythology, related to the notion of myths explaining natural phenomena, was that of Friedrich Max Müller. Müller (1866) viewed mythology as a "disease of language". He believed most Indian, Greek and Roman deities were originally just poetical names for forces of nature which over time began to assume divine personalities never even contemplated by their creators. Müller (1866) identifies similarities between the linguistic name of father-god figures from a variety of cultures. Tolkien (1947), however, argues that this disease of language view "can be abandoned without regret" and believes it is more likely that European languages are a disease of mythology!

The relationship between myth and language are studied in far more detail in the structuralist approach to mythology. Lévi-Strauss (1955) remarks that myth has many similarities to language and that it is, in essence, a language "functioning on an especially high level". The structuralist approach addresses all the constituents of a myth. The myth is broken down into *mythemes* which are the smallest component (or units) of a myth. Meaning is derived from the relations between pairs of mythemes, often opposites of one another (e.g. light versus dark, life versus death, etc.). However, Morford & Lenardon (1999) argue that the binary functioning of the human mind or human society may be common, but it is not universal or even necessary in myth.

A modern approach, related to the idea of mythological characters personifying natural phenomena, involves searching for archaeological, geological or astronomical evidence to support the events portrayed in a myth. For example, a dragon in a culture's myth could correspond to volcanic activity that occurred in that culture's past (Barber & Barber 2006), or a comet could account for elaborate descriptions of gods or heroes (McCafferty & Baillie 2005). Witzel (2012) however is critical of this approach, which he deems as over-simplistic, and describes it as looking for a "single method that would illuminate what myth is all about".

An entirely different method is the psychological approach inspired by the works of Sigmund Freud and Carl Jung. Joseph Campbell, for example, was strongly influenced by ideas from psychology. The psychological approach argues that the common motifs and themes that appear in many myths are due to universal subconscious *archetypes* in every person's mind. The idea of the *monomyth* is almost an extension of this (Campbell 1949). The monomyth explains the journey of the hero in many myths often which reflect common human struggles and concerns. Dundes (1984) argues that the Jung archetypes ideology and the Campbell universality approaches are so general, that it is unsurprising they are so widespread.

There are many other comparative approaches to mythology. Lyle (2012), for example, looks for similar structures and roles for gods in many societies. Another approach is that of Witzel (2012), this method identifies common features from a variety of world mythologies that reflect the myths of our earliest ancestors from Africa. Many approaches to comparative mythology aim to find an *Ur-myth*, the idea that all myth evolved from an original single myth.

A recent approach to the study of mythology is a phylogenetical analysis of myths. Phylogenetics is the study of evolutionary relationships among organisms based on parallelism of genes. D'Huy (2013) observes that there are many similarities between biological and mythological evolution. In mythology, instead of genes, mythemes are transmitted. Then evolutionary relatedness between myths can be observed and inferences about human migration and cultural diffusion can be made. A similar approach has been applied to analyse cross-cultural relationships among folktales (Tehrani 2013). These are examples of new exploratory research that open new areas of inquiry.

Here an alternative approach to the study of mythology is presented using the techniques of social-network analysis as described in the previous two chapters. This is not the first network analysis approach to mythology. Choi & Kim (2007) look at a directed network of Greek and Roman myths creating edges between characters who are mentioned in another character's entry. They find that this network follows a power-law degree distribution and has similar properties to the social network of movie actors. As this looks at a dictionary rather than specific myths, the edges will only be a subset of the total edges from various myths. Similarly, for social networks it is better if the edges are undirected as the acquaintanceship should go both ways, as discussed in the previous chapter. The approach, here delving into each particular myth, is far more rigorous.

Initial investigations using these methods have been well received by both the physics (Mac Carron & Kenna 2012) and the humanities communities (Mac Carron & Kenna 2013b). In a report to HERA (Humanities European Research Area), Hazelkorn *et al.* (2013) state that "the work of Mac Carron & Kenna (2012) provides an example of how digital techniques and technologies, applied to literary works, can open-up a new field of inquiry."

## 4.2 Mythological Networks

Social networks from the myths of a variety of cultures are constructed here. To gather the data, a character's name, gender and page number corresponding to their first appearance, are recorded. As shown in chapter 3, different mechanisms for creating edges lead to different network properties. Therefore it is important when gathering the data to have a consistent mechanism for creating edges. In light of this, two distinct types of edges are defined, *friendly* (or social) edges and *hostile* edges.

A friendly edge is created between two characters if:

– they speak directly to one another
– they are in the same immediate family (i.e. siblings or parents)
– they both know one another
– they are present together in a small congregation.

A hostile edge is made between characters if:

– they physically fight against one another
– they are at war with one another.

The edge must go both ways, if one character mentions another this is only a link if it is clear the other also knows the first. Characters can have both friendly and hostile edges. If two characters argue or are always aggressive to one another, this is not a hostile edge. Hostile edges mostly occur where two characters meet on the battlefield but don't interact otherwise.

Longer myths are broken down into smaller sections (generally corresponding to chapter breaks). A new network is started for each section. This allows us to give a weight to the edges based on how many sections two characters interact in.

## 4.3 Data

As mentioned above, a myth here refers to a traditional story concerning a specific group of people. The narratives used for different cultures are laid out in the following sections.

### 4.3.1 Irish Mythology

Irish myths are the best preserved of Celtic mythology (MacKillop 2004). However, a lot of material was either destroyed or possibly never even committed to writing. Of the myths that survived, these were transmitted orally for centuries before being recorded in Christian times (O'Rahilly 1946). Due to religious influences of the Christian monks recording them, it is likely that the pagan aspects of these myths were greatly toned down (Murphy 1961).

Irish mythology is generally divided into four distinct cycles; the *Mythological Cycle*, the *Ulster Cycle*, the *Fenian Cycle* and the *King's Cycle* (Gantz 1981). Narratives from the first three of these are chosen in this study. There tends to be an overlap of characters from the Mythological Cycle in the both the Ulster and Fenian Cycles but this is more prominent with the Ulster Cycle.

When using translations for this study, a persistent issue is that there does not appear to be one consistent convention for the spelling of characters' names. For example Fróech, the hero of *The Cattle Raid of Fróech* is also spelled "Fráech", "Fraích" and "Fraoch". Therefore the same translators are used as much as possible. This makes it difficult to merge all the narratives to get a large network for one specific cycle.

**The Mythological Cycle**

The Mythological Cycle has the least number of surviving texts. These myths tell of the Tuatha Dé Danann (a divine race often identified as the Irish gods (Murphy 1961; O'Rahilly 1946)) and how they came to Ireland. Of these, network data are gathered from the *Second Battle of Mag Tuired* (translations by Elizabeth Gray (CELT 2004a) and Whitley Stokes (CELT 2004b)), the *Wooing of Étaín*, the *Destruction of Da Derga's Hostel* and the *Dream of Óengus* (all translated by Gantz (1981)).

The *Second Battle of Mag Tuired* survives in a 16th century manuscript though it is believed to have been compiled from 9th and 12th century material (Murphy 1961). It deals with how the Tuatha Dé Danann, led by Lugh, defeat their oppressors the Fomorians. This story also alludes to the first battle in the same place where the Tuatha Dé Danann overthrew the previous occupants of Ireland named the Fir Bolg. This narrative effectively lists the entire Irish pantheon detailing each character's magical ability.

The *Wooing of Étaín* is a short story completely preserved in a 15th century manuscript and partially preserved in a 12th century manuscript (Gantz 1981). It is written in a language dated to the 8th or 9th century (MacKillop 2004). It tells of how Midir, of the Tuatha Dé Danann, falls in love with the daughter of the king of Ulster, Étaín. Out of jealousy Midir's wife casts a series of spells that leads to Étaín being reborn to the wife of an Ulster warrior. This time she is married to the High King of Ireland before Midir finds her. Midir steals Étaín and when High King attacks him, Midir tricks him into taking back his own daughter thinking it is his wife.

The *Dream of Óengus* is related to first part of the story the *Wooing of Étaín* dealing with Midir's foster son Óengus. Though it comes from a relatively late source (15th century), it is also mentioned in a 12th century manuscript (Gantz 1981). In the story, Óengus dreams about a girl resulting in him falling in love with her. After he finds her, it is revealed that she turns into a swan every year. Her father will only allow him to marry her if he can identify her as a swan. Óengus correctly identifies her and then turns himself into a swan and they fly away singing beautiful music.

The *Destruction of Da Derga's Hostel* follows on from *The Wooing of Étaín*. It is found in three recensions, the earliest of which is from the 12th century. It details the death of the High King Conare Már, the grandson (or great-grandson) of Étain. Conare's three foster brothers whom he had exiled to Alba (Scotland) attacked the hostel he was staying in with a large band of raiders. Conare is killed in the attack though the hostel is not actually destroyed.

Parts of both the *Wooing of Étaín* and the *Dream of Óengus* are set during the Ulster Cycle. This allows us to merge their networks with those of the Ulster Cycle to create a larger network to analyse.

**The Ulster Cycle**

The Ulster Cycle tells of the heroes of Ireland's northernmost province. It is mostly set during the reign of Conchobar Mac Nessa and frequently deals with Ulster's capricious relationship with its neighbouring province Connacht under the rule of queen Medb. Networks of the following Ulster Cycle narratives are constructed: *Táin Bó Cúailnge* (translated by Kinsella (1969)), the *Cattle Raid of Fróech*, the *Wasting Sickness of Cú Chulaind*, the *Tale of Macc Da Thó's Pig*, the *Intoxication of the Ulaid* and *Bricriu's Feast* (all of which are translated by Gantz (1981)).

The *Táin Bó Cúailnge* (*"Táin"* from here on, also known as the Cattle Raid of Cooley) is one of Ireland's greatest epics. It survives in three manuscripts from the 12th to 14th centuries. It describes the invasion of Ulster by the armies of Connacht in an attempt to obtain a magical bull. The men of Ulster are under a spell making them unable to fight so the boy-hero Cúchulainn steps up to defend Ulster by invoking the right of single combat once a day. This gives enough time for the men of Ulster to recover and gather their army. Connacht, however, succeeds in obtaining the bull.

The *Táin* is accompanied by a number of remscéla (pre-tales) describing the protagonists' backgrounds. Among these are *How Conchobar was Begotten, The Pangs of Ulster, The Exile of the sons of Uiliu, How Cúchulainn was Begotten, Cúchulainn's Courtship of Emer and his Training in Arms with Scáthach, The Death of Aife's One Son* and *The Quarrel of the two Pig-keepers and how the Bulls were Begotten*. These are all analysed with the *Táin* network and not separately.

The *Cattle Raid of Fróech* (or *Táin Bó Fróech*) contains another mixture of Mythological Cycle characters and Ulster Cycle characters. It is found in a number of manuscripts, the earliest from the 12th century but it is believed to be an 8th century narrative (MacKillop 2004). The protagonist, Fróech, falls in love with Findabair, the daughter of Ailill and Medb of Connacht. This causes concern for Ailill and Medb as they promised Findabair to a king of Ireland. They plan to have Fróech killed by a water monster but, with the help of Findabair, he survives and kills the monster. Ailill and Medb next plan to have Findabair killed but Fróech outsmarts them and so they allow them to be together. Later in the story Findabair and Fróech's cattle are stolen. With the help of the Ulster hero Conall Cernach they travel to Alba and then the Alps and recover them. Fróech is killed by Cúchulainn in the *Táin*.

The *Wasting Sickness of Cú Chulaind* (spelled Cúchulainn in the *Táin*) is a combination of a 10th and an 11th century tale found in a 12th century manuscript (MacKillop 2004). The narrative begins with an account of how the people of Ulster celebrate Samhain (Halloween). Cúchulainnn is put to sleep after trying to hunt a flock of birds. In his dreams two women abuse him and he remains prostate for almost a year. He is nursed by a woman named Fand whom he falls in love with. His wife Emer is jealous and she and Fand fight for Cúchulainn's love. Fand eventually yields. After they Cúchulainn and Emer drink a vial of forgetfulness.

The *Tale of Macc Da Thó's Pig* is another popular tale surviving in at least six manuscripts, the earliest of which is from the 12th century. This story includes almost every major Ulster Cycle figure except Cúchulainn. Once again Ulster and Connacht are in dispute, this time over a magical dog instead of a bull. The owner of the hound, Macc Da Thó is troubled by both provinces demanding his hound so devises a plan to give the hound to both parties and let them fight for it. He invites them both to his hostel and feeds them with an enormous pig. Eventually a fight breaks out and Macc Da Thó unleashes the hound to see which side it chooses. It chooses Ulster but is killed while routing the Connacht men. Gantz (1981) views this narrative as a parody of the *Táin*.

The *Intoxication of the Ulaid* is found incomplete in two 12th century manuscripts. It is one of the more bizarre Irish narratives with no definite plot. The heroes of Ulster get drunk and lose their way on their chariots ending up in Munster instead of Ulster. They become trapped in an iron house surrounded by their enemies. The end of the story is contained in one of the versions however it is not clear how they escape the trap.

*Bricriu's Feast* is an 8th century tale found in various manuscripts, the earliest being from the 12th century. At Bricriu's feast, three Ulster heroes compete for the champion's portion – the largest serving of food. To decide who gets it they are given a series of trials all of which Cúchulainn wins.

**The Fenian Cycle**

The Fenian Cycle is set after the Ulster Cycle and tells the tales of the hero Fionn Mac Cumhaill and his band of warriors the Fianna. The texts used to gather network data are *Tales of the Elders of Ireland* (translated by Dooley & Roe (1999)) and *Fianaigecht* (tranlsated by Meyer (1910)).

*The Colloquy of the Ancients* (also known as *Tales of the Elders of Ireland*) collects Fenian stories and poetry from four manuscripts. The work is normally dated to the late 12th or early 13th century. The story is framed by the warriors Caílte and Fionn's son, Óisín, who lived until Christian times, reciting tales of the Fianna to St Patrick as they travel around Ireland. Many of the stories deal with the rivalry between their leader Fionn Mac Cumhaill and his tenuous relationship with Goll Mac Morna. Members of the Tuatha Dé Danann also crop up in numerous stories. Like the Táin, the *The Colloquy of the Ancients* has a heavy emphasis on place-naming (Kinsella 1969; Dooley & Roe 1999).

*Fianaigecht* is a compilation of the earliest accounts of Fionn and his Fianna. These are six texts and poems dating from the 7th up the 14th century (Meyer 1910). Unlike *The Colloquy of the Ancients*, these tales are not narrated by an ageing warrior of the Fianna to St Patrick. It contains six narratives about various members of the Fianna and ends with a different account of the death of Fionn.

### 4.3.2 Greek & Roman Mythology

Classical mythology is the body of myths from the ancient Greeks and Romans. Greek myths tend to concern their gods and heroes. Roman mythology took a more historical rather than religious approach. As the Romans were influenced by Greek myths, some stories and events tended to be rationalised and reinterpreted (Morford & Lenardon 1999). The texts used to create networks are Homer's *Iliad* (translated by E.V. Rieu (2003)) and *Odyssey* (translated by Kirk, G.S. (1980)) and Virgil's *Aeneid* (translated by Oakley, M. (2004)).

The *Iliad* is an epic poem attributed to Homer and is dated to the 8th century BC (E.V. Rieu 2003). It takes place over a period of a few weeks from the final year of the Trojan War. During this time the Greek King Agamemnon quarrels with his greatest warrior Achilles resulting in the latter's refusal to fight. After the Trojan hero Hector kills Achilles' brother-in-arms, Achilles is enraged and rejoins the battle. The story ends shortly after Hector is killed in single combat by Achilles.

The *Odyssey* is effectively a sequel to the *Iliad*. It is also attributed to Homer and dated to the late 8th century BC. It is set ten years after the Trojan War. Odysseus still has not returned to Ithaca and his wife, Penelope, is under pressure to remarry and has many suitors. Odysseus, however, has been captured by the nymph Calypso. He is kept prisoner on an island for seven years until the gods allow him to leave. When he returns to Ithaca he disguises himself as a beggar and outperforms Penelope's suitors in a competition for her hand. With the help of his son he then kills all the suitors and reveals his identity to Penelope.

*The Aeneid* is an epic Latin poem composed by Virgil between 29 and 19 BC. The story deals with the protagonist Aeneas and his destiny to begin the Roman empire. After the events of the Trojan war, a group of Trojans led by Aeneas escape. They land in Carthage where the queen, Dido, falls in love with Aeneas (with the help of the gods). After a year they leave and eventually settle in Latium. Some of the locals, lead by Turnus, engage in a series of battles. The story culminates with Aeneas defeating Turnus in single combat.

### 4.3.3 Germanic Mythology

Germanic mythology refers to the myths of the Germans, Scandinavians and Anglo-Saxons (Leeming 2005). Similar to Irish mythology mentioned earlier, most of the material was committed to writing after Christianisation. The material was mostly passed orally before this and some myths have been distorted by the Christian monks recording it (Lindow 2002). The texts used here are *Beowulf* (translations by Heaney (1999) and Crossley-Holland (1999)), the *Poetic Edda* (translated by Larrington (1999)), Snorri Sturluson's *Prose Edda* (translated by Byock (2005)), *Völsungasaga* (translated by Byock (1999)), *Nibelungenlied* (translated by Hatto (2004)) and *Orkneyinga saga* (tanslated by Pálsson & Edwards (1981)).

*Beowulf* is an Old English heroic epic, set in Scandinavia. A single codex survives which is dated from between the 8th and early 11th centuries (Bloomfield 1970). In the poem, Beowulf, a Gaetish hero, vanquishes Grendel, a monster who had been attacking the mead hall of the Danes. After returning home Beowulf eventually becomes the king of his people. Fifty years after his battle with Grendel, his kingdom is attacked by a dragon. Beowulf battles it and, although he defeats it, he is fatally wounded in the process.

The *Poetic Edda* is a collection of Old Norse poems mostly preserved in a 13th century Icelandic manuscript. Many of the poems pre-date the conversion of Scandinavia to Christianity, allowing for a glimpse into the Norse religion (Larrington 1999). The poems tell of the world's creation from the body of the frost giant Ymir to its fiery destruction with the aid of the giant Surt. It also contains many Germanic heroic legends.

The *Prose Edda* is a 13th century compilation of tales attributed to Snorri Stulsuson. It contains many of the same stories as the *Poetic Edda* and cites it as a source (Byock 2005). A distinct difference is it contains an euhemerised Christian account of the origins of the Norse gods. Similar to the *Aeneid*, they are described as warriors who left Troy after the fall of Trojan War, eventually settling in northern Europe.

*Völsungasaga* (or *Saga of the Völsungs*) was written by an unknown Icelandic author in the 13th century (Byock 1999). It is based on older Norse myths and a version of it is found in the both *Poetic* and *Prose Edda*s. The tale follows the descendants of Volsung, initially his son Sigmund followed by his son, Sigurd "the dragon slayer". Sigurd is in love with the Valkyrie Brynhild, however, he is given a potion that makes him forget his love for Brynhild and then marries Gudrun. After Sigurd is killed by Gudrun's brothers the tale follows Gudrun and her family as they quarrel with other kings. Nearly all the characters in this section can be identified with historical characters (Byock 1999).

The *Nibelungenlied* (or the *Song of the Nibelungs*) is an epic written in the early 13th century by an unnamed poet (Hatto 2004). Similar to many of the other epics it has an oral tradition far pre-dating its time of writing. There are distinct parallels between the characters Siegfried and Kriemhild to the *Völsungasaga*'s Sigurd and Gudrun (whom also appear in the two *Edda*s) (Leeming 2005). Siegfried wins queen Brunhild for King Gunther in exchange for Gunther's sister Kriemhild. However Siegfried is murdered by Gunther's vassals and his treasure is stolen. Kriemhild avenges her husband by killing his murderers but is killed in the process.

*Orkneyinga saga* is a history of the earls of the Orkney Islands from the 9th until the 12th century. It was written by an unknown Icelandic author in the 13th century. It is both Norse and Scottish and has similarities in its narrative style to the sagas of the Icelanders combining elements of fiction and history (Pálsson & Edwards 1981). It details wars, vengeance and reconciliation between the earls and their involvement in the complexities of the old and the new religions.

### *4.3.4 Sagas of Icelanders*

The *Íslendinga sögur*, or Sagas of Icelanders, are texts describing events purported to have occurred in Iceland in the period following its settlement in late 9th to the early 11th centuries. It is generally believed that the texts were written in the 13th and 14th centuries by authors of unknown or uncertain identities but they may have oral prehistory (O'Donoghue 2004). The texts focus on family histories and genealogies and reflect struggles and conflicts amongst the early settlers of Iceland and their descendants. The sagas describe many events in clear and plausible detail and are considered to be amongst the gems of world literature and cultural inheritance.

Five of the sagas contain large casts of characters and are selected for individual study here. These are *Gísla saga Súrssonar* (*Gisli Sursson's Saga*), *Vatnsdæla saga* (*Saga of the People of Vatnsdal*), *Egils saga Skallagrímssonar* (*Egil's Saga*), *Laxdæla saga* (*Saga of the People of Laxardal*) (all of which are contained in Smiley 2000) and *Njáls saga* (translated by Bayerschmidt & Hollander 1998). Smiley (2000) also contains 13 shorter sagas and tales, many of which contain recurring characters. These are merged with the five larger sagas and referred to as "18 Sagas" in the analysis. These are *Bolli Bollason's Tale*, the *Saga of Hrafnkel Frey's Godi*, the *Saga of the Confederates*, the *Saga of Gunnlaug Serpent Tongue*, the *Saga of Ref the Sly*, the *Saga of the Greenlanders*, *Eirik the Red's Saga*, *The Tale of Thorstein Staff-Struck*, *The Tale of Halldor Snorrason II*, *The Tale of Sarcastic Halli*, *The Tale of Thorstein Shiver*, *The Tale of Audun from the West Fjords*, and *The Tale of the Story-wise Icelander*.

*Gísla Saga* is an outlaw narrative centred on human struggles, as the eponymous character is "on the run" for 13 years before being finally killed. It is set in the period 940-980 AD. There are two versions of this and we use the version translated by Regal in (Smiley 2000). *Gísla Saga*, is an "outlaw saga" is mostly centred on one character rather than on a society and in this sense it is quite different to the other sagas considered here.

*Vatnsdæla Saga* is essentially a family chronicle. It follows the settling of Ingimund, the grandson of a Norwegian chieftain, in Iceland with his family until the arrival of Christianity in the late 10th century.

*Egils Saga* tells of the exploits of a warrior-poet and adventurer. The story begins in Norway with Egil's grandfather and his two sons. After one of them is killed, as a result of a dispute with the king, the family leaves to settle in Iceland. The latter part of the story is about the life of Egil himself. *Egils Saga* is also noteworthy in that a significant proportion of it is set outside Iceland. It begins in Norway with the protagonist's family, where about a third of the saga's characters first appear. Later in the story Egil travels to Norway, amongst other places. Therefore the network contains overlapping social structures rather than a single coherent one.

*Laxdæla Saga* tells of the people of an area of western Iceland from the late 9th to the early 11th century. It has the second highest number of preserved medieval manuscripts and also contains the second largest network.

*Njáls saga* is widely regarded as the greatest piece of prose literature of Iceland in the Middle Ages and more vellum manuscripts containing it have survived compared to any other saga (Mahnusson & Pálsson 1960). It also contains the largest saga-society network. The epic deals with blood feuds, recounting how minor slights in the society could escalate into major incidents and bloodshed. The events described are purported to take place between 960 and 1020 AD and, while most archaeologists believe the major occurrences described in the saga to be historically based, there are clear elements of artistic embellishment.

### 4.3.5 Welsh & Arthurian Mythology

Welsh mythology comes from a variety of sources, only some of which contain the mythology of pre-Christian Britain. These are collected in the first four "branches" of the *Mabinogion* (Davies 2007). The rest often concerns Arthurian romances and legends, the earliest of which are found in two ninth century manuscripts (Leeming 2005). The texts used here are the *Mabinogion* (translated by Davies (2007)), Thomas Malory's *Le Morte d'Arthur*, Chrétien de Troyes' *Arthurian Romances* (translated by De Troyes (1991)), *Queste del Saint Graal* (translated by Matarasso (1969)) and Gottfried von Strassburg's *Tristan*.

The *Mabinogion* is a collection of eleven prose stories found in two medieval manuscripts from the 14th century. The tales were not conceived as a single collection and the link between them is often tenuous at best (Davies 2007). As with many of the narratives employed here, they had a long oral tradition before being committed to writing. There are two distinct sections in the *Mabinogion*. The first four contain tales known as "the four branches of the *Mabinogion*". These are of a more mythological nature, containing characters with supernatural abilities. Of the remaining narratives, five of them are set in Arthurian times and the remaining two tell of early British history.

*Le Morte d'Arthur* is a compilation by Thomas Malory of chivalric romances about King Arthur and his knights. Malory interprets French and English tales about these characters as well as adding some original content. It is set in the 5th century, mostly in Britain and France. There are eight tales, beginning with the birth and rise of Arthur, and ending with his death and the breaking of the Knights of the Round Table.

*Arthurian Romances* is a collection of all five of Chréten de Troyes' chivalric romances. De Troyes drew inspiration from both Breton minstrels and Geoffrey of Monmouth's 12th century pseudo-historical *Historia Regum Britanniæ* (History of the Kings of Britain). These are the earliest of the romances and they added many of the staples of Arthurian literature, for example the affair of Sir Lancelot and Lady Guinevere, the Holy Grail and the castle of Camelot. Each tale centres on the adventures of a different knight, Eric, Cligès, Yvain, Lancelot and Perceval.

*Queste del Saint Graal* (or the *Quest for the Holy Grail*) is part of the of a compilation known as the *Prose Lancelot*. It is thought to have been written in the 13th century and is one of the sources Thomas Malory used (Matarasso (1969)). The story focuses on five knights; Lancelot, Perceval, Bors, Gawain and Galahad and their search for the Holy Grail. Galahad, Lancelot's son, succeeds in finding it and, after obtaining it, ascends into heaven. After this the grail is also brought to heaven never to be seen by man again.

*Tristan* or *Tristan and Iseult* is a chivalric romance pre-dating the Arthurian romances above. The first written version was composed by Thomas of Britain in the 12th century, however it is believed there are earlier versions which are now lost (von Strassburg 2004). Only fragments of Thomas' poem survive but it was used by Gottfried von Strassburg as a source in the early 13th century. The story details how the young prince Tristan wins the hand of Isolde for his uncle and king, Mark. However, they accidentally drink a love potion intended for the king and his bride and begin an illicit affair. Isolde marries Mark but continues her affair with Tristan. When Mark finds out, Tristan flees to Normandy, where he meets another Isolde, Isolde of the White Hands, and finds himself conflicted over which Isolde he loves.

### *4.3.6 World Mythology*

Network data for three non-European myths are also gathered. These are the *Epic of Gilgamesh* (translated by George (2002)), the *Popol Vuh* and *Navaho Indian Myths*.

The *Epic of Gilgamesh* is the world's oldest epic and is among the earliest surviving works of literature. Fragments of it have been discovered on many tablets providing a number of different versions. The oldest fragments date from the 18th century BC. The most complete version, known as the "standard Babylonian" version dates from the 13th to the 10th century BC. The first part of the story deals with the friendship and adventures of King Gilgamesh and Enkidu. After Enkidu dies, Gilgamesh undertakes a journey in an attempt to discover the secret of eternal life which he never finds.

The *Popol Vuh* is a collection of myths from the K'iche' people in Central America. They were recorded in Spanish in the early 18th century by Francisco Ximénez, a Dominican friar. The protagonist of many of the epic tales are the twins Hunahpú and Xbalanqué. The *Popol Vuh* also features a creation myth and has a strong emphasis on cosmology.

*Navaho Indian Myths* are a collection of tales from the Navajo (or Navaho) people recorded in 1928 (O'Bryan *et al.* 1993). An elderly chief approached O'Bryan asking her to record the legends for future generations of his people. These narratives also include a creation myth and show the importance of omens to the Navajo people. The character Atse'hastqin, the First Man, is central to the myths.

# 5. NETWORK ANALYSIS

In this chapter, the methodology introduced in chapter 2 is applied to each of the narratives discussed in chapter 4. An overview of the results is presented in the next chapter. The core properties of each network are listed in table 5.1 and table 5.2.

In some cases two translations of a narrative were used. These were the *Second Battle of Mag Tuired*, the *Táin*, Beowulf and the *Popol Vuh*. The edges were found to be almost identical. In most cases, a second translation allows for further clarity when there is uncertainty in creating an edge.

Many of the edges in these networks are both friendly and hostile. When using the community detection algorithm, we are interested in friendly edges as hostile links are often formed between opposing factions. When the term "purely hostile" is used, edges that are both friendly and hostile at different points in the narrative are treated as friendly. In networks which contain relatively few hostile edges, the friendly assortativity and structural balance is not reported.

As well as the mythological networks presented here, two further fictional networks are analysed at the end of the chapter. This allows us to compare the networks of contemporary narratives with mythological ones. These are J.R.R Tolkein's *Lord of the Rings* trilogy and *The Girl with the Dragon Tattoo* by Stieg Larsson (referred to as "Dragon Tattoo" in table 5.4), the first book of the Millenium trilogy, both of which contain a large cast of characters. These are by no means an attempt to characterise the entire corpus of world fiction, just two mere examples to contextualise. There are many other studying the social networks of fiction (e.g. Elson *et al.* (2010), Stiller *et al.* (2003), etc.).

## *5.1 Irish Mythology*

A number of the Irish texts contain less than 50 vertices but many of these characters appear in other texts from the same cycle. These texts are not reported individually in table 5.1 and table 5.2, instead they are merged with other texts from their corresponding cycle as described below.

*Tab. 5.1:* Comparison of the properties of the mythological networks. The number of vertices is denoted by $N$ and the number of edges by $L$. The average degree is $\langle k \rangle$, the maximum degree is $k_{\max}$, and $\langle k^2 \rangle$ is the mean of the square of the degree. The degree assortativity is $r_k$ where an asterisk (*) denotes that the error is larger than the value, $r_C$ is the clustering assortativity and $r_P$ is the Pearson similarity.

| | $N$ | $L$ | $\langle k \rangle$ | $k_{\max}$ | $\langle k^2 \rangle$ | $r_k$ | $r_C$ | $r_P$ |
|---|---|---|---|---|---|---|---|---|
| **Irish** | | | | | | | | |
| *Mag Tuired* | 95 | 332 | 6.99 | 43 | 92.5 | -0.16 | 0.07 | 0.07 |
| *Da Derga's Hostel* | 126 | 410 | 6.51 | 71 | 107.3 | -0.18 | 0.31 | 0.08 |
| Étaín + Óengus | 53 | 108 | 4.08 | 18 | 31.3 | -0.23 | -0.02* | 0.07 |
| *Táin Bó Cúailnge* | 422 | 1266 | 6 | 168 | 243.8 | -0.35 | -0.30 | 0.14 |
| *Intoxication of the Ulaid* | 82 | 194 | 4.73 | 26 | 45.3 | -0.08 | 0.24 | 0.05 |
| *Bricriu's Feast* | 76 | 591 | 15.55 | 39 | 387.2 | 0.48 | 0.40 | 0.10 |
| Ulster Cycle | 229 | 1070 | 9.34 | 81 | 247.2 | -0.03* | 0.19 | 0.06 |
| Mythological + Ulster | 374 | 1574 | 8.42 | 83 | 201.4 | -0.01* | 0.21 | 0.03 |
| *Colloquy of the Ancients* | 732 | 2268 | 6.20 | 159 | 205.8 | -0.01* | 0.48 | 0.02 |
| *Fianaigecht* | 193 | 338 | 3.50 | 53 | 34.5 | -0.07 | 0.47 | 0.02 |
| *Fianaigecht* (no gene) | 138 | 283 | 4.10 | 53 | 46.6 | -0.14 | 0.30 | 0.05 |
| **Greek & Roman** | | | | | | | | |
| *Iliad* | 694 | 2684 | 7.73 | 106 | 212.4 | -0.08 | 0.53 | 0.02 |
| *Odyssey* | 301 | 1019 | 6.77 | 112 | 126.9 | -0.08 | 0.38 | 0.04 |
| *Aeneid* | 444 | 982 | 4.42 | 118 | 76.2 | -0.13 | 0.33 | 0.03 |
| **Germanic** | | | | | | | | |
| *Beowulf* | 72 | 167 | 4.64 | 27 | 39.9 | -0.12 | -0.05* | 0.04 |
| *Poetic Edda* | 354 | 1138 | 6.43 | 26 | 92.2 | 0.70 | 0.64 | 0.01 |
| *Prose Edda* | 374 | 1955 | 10.45 | 80 | 252.2 | 0.33 | 0.52 | 0.02 |
| *Prose Edda* (no dwarves) | 315 | 1155 | 7.33 | 80 | 137.0 | 0.02* | 0.36 | 0.02 |
| *Völsungasaga* | 103 | 304 | 5.9 | 24 | 57.1 | -0.01* | -0.04* | 0.03 |
| *Nibelungenlied* | 66 | 313 | 9.48 | 43 | 178.2 | -0.28 | -0.22 | 0.23 |
| *Orkneyinga saga* | 441 | 1197 | 5.43 | 73 | 77.7 | -0.09 | 0.20 | 0.02 |
| **Sagas of Icelanders** | | | | | | | | |
| *Gísla saga* | 103 | 254 | 4.93 | 44 | 55.5 | -0.15 | 0.01* | 0.06 |
| *Vatnsdæla saga* | 132 | 290 | 4.39 | 31 | 39.2 | 0.00* | 0.08 | 0.02 |
| *Egils Saga* | 292 | 770 | 5.27 | 59 | 64.3 | -0.07 | 0.28 | 0.02 |
| *Laxdæla saga* | 332 | 894 | 5.39 | 45 | 61.0 | 0.19 | 0.25 | 0.01 |
| *Njáls saga* | 575 | 1612 | 5.61 | 83 | 103.7 | 0.01* | 0.12 | 0.01 |
| 5 Sagas | 1282 | 3729 | 5.82 | 87 | 93.5 | 0.04 | 0.17 | 0.00 |
| 18 Sagas | 1546 | 4267 | 5.52 | 88 | 83.9 | 0.06 | 0.16 | 0.00 |
| **Welsh & Arthurian** | | | | | | | | |
| *Mabinogion* | 666 | 2427 | 7.29 | 135 | 217.5 | 0.19 | 0.37 | 0.02 |
| *Mabinogion* 4 Branches | 75 | 217 | 5.79 | 26 | 57.9 | -0.03* | -0.00* | 0.05 |
| *Mabinogion* 7 Tales | 601 | 2212 | 7.36 | 135 | 232.0 | 0.15 | 0.38 | 0.02 |
| *Le Morte d'Arthur* | 504 | 2497 | 9.91 | 237 | 452.4 | -0.23 | -0.04 | 0.09 |
| *Arthurian Romances* | 255 | 714 | 5.60 | 91 | 122.5 | -0.29 | -0.12 | 0.08 |
| *Queste del Saint Graal* | 122 | 225 | 3.69 | 36 | 44.4 | -0.23 | 0.11 | 0.08 |
| *Tristan* | 49 | 133 | 5.43 | 44 | 74.3 | -0.37 | -0.33 | 0.28 |
| **World** | | | | | | | | |
| *Epic of Gilgamesh* | 46 | 81 | 3.52 | 19 | 27.7 | -0.34 | 0.10* | 0.11 |
| *Popol Vuh* | 98 | 409 | 8.35 | 27 | 123.4 | -0.32 | -0.05* | 0.10 |
| *Navaho Indian Myths* | 140 | 283 | 4.04 | 32 | 35.4 | -0.18 | 0.31 | 0.03 |

*Tab. 5.2:* Comparison of the properties of the mythological networks. The average path length is given by $\ell$ and $\ell_{\mathrm{rand}}$ for its random graph. The diameter is $\ell_{\max}$. The clustering coefficient is denoted by $C$ and the random clustering coefficient is $C_{\mathrm{rand}}$. The transitivity is $C_T$ and $C_{\mathrm{n}}$ is the naïve value for the transitivity. The size of the giant component as a percentage of $N$ is given by $G_C$.

| | $\ell$ | $\ell_{\mathrm{rand}}$ | $\ell_{\max}$ | $C$ | $C_{\mathrm{rand}}$ | $C_T$ | $C_{\mathrm{n}}$ | $G_C$ |
|---|---|---|---|---|---|---|---|---|
| **Irish** | | | | | | | | |
| *Mag Tuired* | 2.64 | 2.55 | 6 | 0.72 | 0.07 | 0.50 | 0.23 | 97.9% |
| *Da Derga's Hostel* | 2.76 | 2.77 | 7 | 0.64 | 0.05 | 0.50 | 0.29 | 98.4% |
| Étaín + Óengus | 2.93 | 2.92 | 6 | 0.49 | 0.08 | 0.30 | 0.21 | 100.0% |
| *Táin Bó Cúailnge* | 2.80 | 3.55 | 8 | 0.73 | 0.01 | 0.10 | 0.62 | 98.8% |
| *Intoxication of the Ulaid* | 3.04 | 2.96 | 8 | 0.44 | 0.06 | 0.38 | 0.19 | 91.5% |
| *Bricriu's Feast* | 2.29 | 1.87 | 5 | 0.70 | 0.20 | 0.84 | 0.48 | 100.0% |
| Ulster Cycle | 3.02 | 2.67 | 10 | 0.60 | 0.04 | 0.52 | 0.30 | 98.3% |
| Mythological + Ulster | 3.42 | 3.01 | 10 | 0.62 | 0.02 | 0.49 | 0.17 | 98.9% |
| *Colloquy of the Ancients* | 3.66 | 3.80 | 13 | 0.38 | 0.01 | 0.51 | 0.23 | 77.5% |
| *Fianaigecht* | 17.48 | 4.24 | 64 | 0.29 | 0.02 | 0.32 | 0.12 | 92.7% |
| *Fianaigecht* (no gene) | 3.41 | 3.58 | 10 | 0.41 | 0.03 | 0.33 | 0.19 | 89.9% |
| **Greek & Roman** | | | | | | | | |
| *Iliad* | 3.49 | 3.42 | 11 | 0.44 | 0.01 | 0.45 | 0.13 | 99.4% |
| *Odyssey* | 3.29 | 3.18 | 8 | 0.45 | 0.02 | 0.38 | 0.15 | 98.3% |
| *Aeneid* | 3.57 | 4.21 | 9 | 0.41 | 0.01 | 0.21 | 0.13 | 88.3% |
| **Germanic** | | | | | | | | |
| *Beowulf* | 2.38 | 2.91 | 6 | 0.57 | 0.06 | 0.37 | 0.17 | 69.4% |
| *Poetic Edda* | 4.99 | 3.34 | 13 | 0.57 | 0.02 | 0.84 | 0.08 | 69.2% |
| *Prose Edda* | 4.36 | 2.78 | 14 | 0.67 | 0.03 | 0.81 | 0.14 | 85.8% |
| *Prose Edda* (no dwarves) | 4.32 | 3.10 | 14 | 0.60 | 0.02 | 0.58 | 0.14 | 86.7% |
| *Völsungasaga* | 3.41 | 2.79 | 7 | 0.68 | 0.06 | 0.44 | 0.12 | 100.0% |
| *Nibelungenlied* | 2.14 | 2.11 | 5 | 0.69 | 0.14 | 0.48 | 0.51 | 97.0% |
| *Orkneyinga saga* | 5.04 | 3.76 | 21 | 0.50 | 0.01 | 0.27 | 0.07 | 99.5% |
| **Sagas of Icelanders** | | | | | | | | |
| *Gísla saga* | 3.38 | 3.04 | 11 | 0.60 | 0.05 | 0.26 | 0.21 | 98.1% |
| *Vatnsdæla saga* | 3.86 | 3.41 | 10 | 0.49 | 0.03 | 0.34 | 0.11 | 97.0% |
| *Egils Saga* | 4.19 | 3.57 | 12 | 0.56 | 0.02 | 0.38 | 0.08 | 97.3% |
| *Laxdæla saga* | 5.01 | 3.60 | 16 | 0.45 | 0.02 | 0.41 | 0.06 | 99.1% |
| *Njáls saga* | 5.14 | 3.85 | 24 | 0.42 | 0.01 | 0.26 | 0.09 | 100.0% |
| 5 Sagas | 5.10 | 4.24 | 16 | 0.46 | 0.00 | 0.28 | 0.03 | 98.8% |
| 18 Sagas | 5.58 | 4.46 | 19 | 0.46 | 0.00 | 0.28 | 0.02 | 98.6% |
| **Welsh & Arthurian** | | | | | | | | |
| *Mabinogion* | 3.83 | 3.48 | 11 | 0.48 | 0.01 | 0.63 | 0.17 | 76.0% |
| *Mabinogion* 4 Branches | 3.17 | 2.63 | 8 | 0.52 | 0.08 | 0.46 | 0.19 | 100.0% |
| *Mabinogion* 7 Tales | 3.52 | 3.42 | 10 | 0.47 | 0.01 | 0.64 | 0.21 | 72.9% |
| *Le Morte d'Arthur* | 2.76 | 2.96 | 7 | 0.59 | 0.02 | 0.27 | 0.40 | 98.2% |
| *Arthurian Romances* | 2.92 | 3.38 | 6 | 0.67 | 0.02 | 0.19 | 0.30 | 99.2% |
| *Queste del Saint Graal* | 3.35 | 3.74 | 13 | 0.45 | 0.03 | 0.22 | 0.27 | 90.2% |
| *Tristan* | 1.99 | 2.46 | 4 | 0.75 | 0.11 | 0.27 | 0.61 | 100.0% |
| **World** | | | | | | | | |
| *Epic of Gilgamesh* | 2.54 | 3.08 | 5 | 0.46 | 0.08 | 0.27 | 0.29 | 93.5% |
| *Popol Vuh* | 2.80 | 2.39 | 6 | 0.55 | 0.09 | 0.42 | 0.23 | 94.9% |
| *Navaho Indian Myths* | 3.81 | 3.62 | 9 | 0.44 | 0.03 | 0.26 | 0.11 | 92.1% |

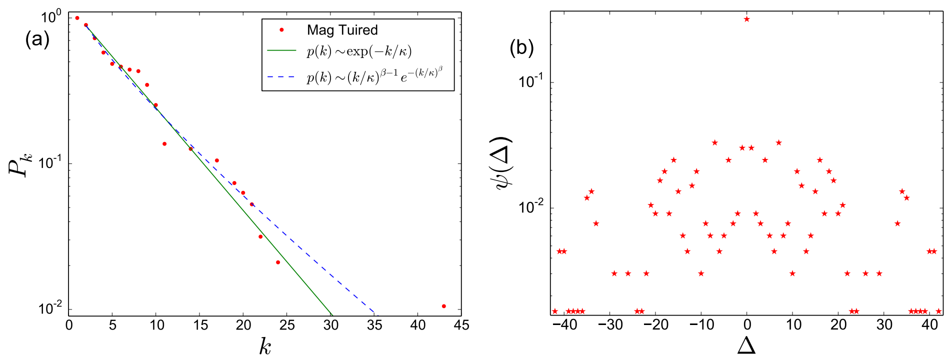



*Fig. 5.1:* Panel (a) shows the complementary cumulative degree distribution for the *Second Battle of Mag Tuired* fitted with an exponential distribution (continuous green line) and a Weibull distribution (dashed blue line). Panel (b) shows the $\psi(\Delta)$ distribution. With the exception of $\Delta = 0$, there is almost equal probability of having a large or low value of $\Delta$.

With the exception of *Bricriu's Feast*, most Irish mythological networks are small world, disassortative, have low positive Pearson similarity measures and their clustering coefficients are not much larger than the naïve prediction. They are also structurally balanced, however only the *Táin* contains an abundance of hostile links. These networks have different properties to the general social networks in Chapter 3. The specifics of each network are discussed below.

### *5.1.1* Second Battle of Mag Tuired

The *Second Battle of Mag Tuired* (referred to as *Mag Tuired* in table 5.1 and table 5.2) contains 95 characters, each of which interacts with almost 7 others on average. Its degree distribution is shown in fig. 5.1 (a). The most likely candidate model is a Weibull distribution. This has parameters $\beta = 0.83 \pm 0.09$ and $\kappa = 5.04 \pm 0.03$ and is represented by the dashed blue line. The AIC$_c$ and BIC weights also give strong support for an exponential distribution however. This has $\kappa = 6.18 \pm 0.05$ and is represented by the continuous green line. We observe that this fits all the data well except for the highest degree vertex.

The network is disassortative with $r_k = -0.16 \pm 0.05$. The protagonist, Lugh, has a degree $k = 43$ which is almost twice that of the next highest character, Nuada Airgetlám, $k = 23$ who was the king of Ireland in certain parts of the story. This large deviation in degree, and Lugh's frequent encounters with low degree vertices are the main cause of the disassortativity. This is shown visually by the frequent large values of $\Delta$ in the $\psi(\Delta)$ distribution in fig. 5.1 (b). The network is mildly clustering assortative, $r_c = 0.07 \pm 0.06$, and has a positive Pearson similarity measure, $r_P = 0.07$, indicating vertices share the same neighbours.

The *Second Battle of Mag Tuired* is small world with an average path length of $\ell = 2.64$ and a clustering coefficient of $C = 0.72$. The naïve value for the transitivity, $C_{\mathrm{n}} = 0.23$, is not much smaller than actual value, $C_T = 0.50$. The network is fragile to the targeted removal of vertices but robust to their random removal. Removing the 5 characters with highest betweenness leaves the giant components 57% of its original size. However, upon the random removal of one quarter of the vertices (24 characters), averaging over 30 realisations, the giant component still contains 87% of the nodes.

The network contains 46 cliques with an average size of $\langle C_s \rangle = 4.3$. The largest clique contains 11 vertices. The Girvan-Newman algorithm finds 6 communities with modularity $Q = 0.59$. This separates the two factions in the battle and it finds sub-communities within one faction. Although this narrative tells of a battle, the two factions are not exclusively hostile to one another as most characters who fight also had a friendly interaction at some point in the tale. There are only 10 purely hostile edges in the network. For this reason the community structure of the friendly network is similar to that of the overall one. For the friendly network, the Girvan-Newman algorithm finds 5 communities with a modularity of $Q = 0.59$ and again separates the two factions who are in the battle.

### *5.1.2* Destruction of Da Derga's Hostel

The *Destruction of Da Derga's Hostel* is the largest of the four Mythological Cycle networks with 126 characters. The protagonist, the high king Conare Már interacts with 71 of these. The next highest degree is $k = 25$. As a result of this, the degree distribution (fig. 5.2 (a)) is similar to that the *Second Battle of Mag Tuired*, in that it is well fitted by an exponential ($\kappa = 5.99 \pm 0.03$) except for the vertex with the highest degree. Using the entire distribution however, it is best fitted by a log-normal with parameters $\mu = 1.32 \pm 0.07$ and $\sigma = 1.08 \pm 0.05$. This is has a longer tail than the Weibull distribution of the *Second Battle of Mag Tuired* as $k_{\max}$ is much larger.

This network is also small world and disassortative with $r_k = -0.18 \pm 0.03$. As shown at the limits of fig. 5.2 (b), the disassortativity is due to the interactions of the highest degree character with many low-degree characters. The network has a high transitivity, $C_T = 0.50$, but its naïve estimate is also relatively high, $C_{\mathrm{n}} = 0.29$.

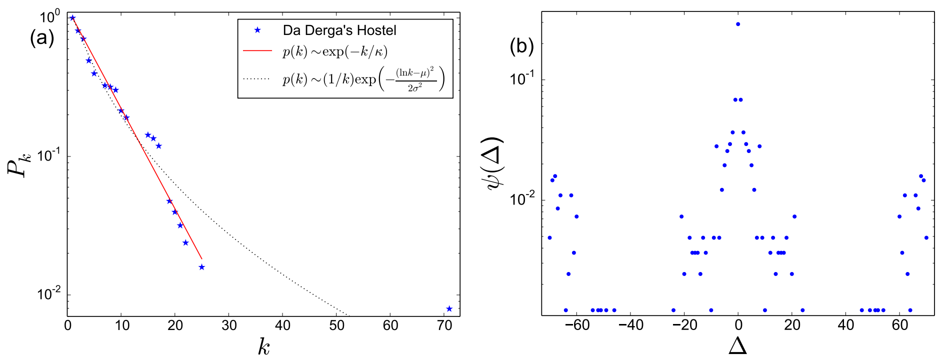



*Fig. 5.2:* Panel (a) shows the complementary cumulative degree distribution for the *Destruction of Da Derga's Hostel* fitted with an exponential distribution (continuous red line) except for the vertex with highest degree. A full fit is made using a log normal distribution (dotted black line). Panel (b) shows the $\psi(\Delta)$ distribution. There are differences in degree at the ends where the king interacts with many low degree characters.

Testing for the robustness, the network is very dependent on the highest degree vertex. Upon removal, the size of the giant component is reduced to 48% of its original size. On the other hand, it is robust to the random removal of vertices.

The largest clique contains 16 vertices and the average size of the 38 cliques is $\langle C_s \rangle = 4.87$. The modularity reaches a plateau at $Q = 0.46$ with 2 communities. These communities comprise the king's men (in which he is central) and most of his enemies. Though this network has many similarities to that of the *Second Battle of Mag Tuired*, it is more dependent on its protagonist. The communities here are centred on their king, unlike the previous network.

### *5.1.3* Wooing of Étaín *and* Dream of Óengus

The *Wooing of Étaín* and the *Dream of Óengus* both contain less than 50 characters. These tales share many of the same characters and their amalgamation contains 53 unique vertices. Due to this small size, the two networks are not treated individually but amalgamated and referred to as the "Étaín + Óengus" in table 5.1 and table 5.2.

As there is only a small number of vertices, it is difficult to decipher the nature of the degree distribution. A log-normal distribution and a Weibull distribution both have strong support from the $AIC_c$ and BIC weights and are shown in fig. 5.3 (a). The network is disassortative with $r_k = -0.23 \pm 0.09$. This can be seen visually

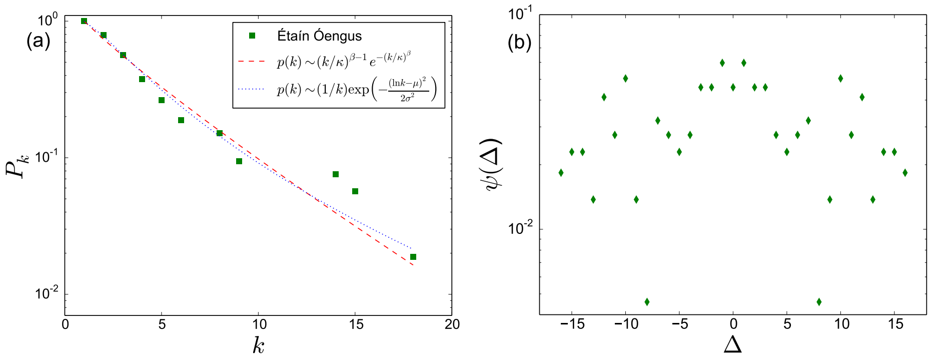



*Fig. 5.3:* Panel (a) shows the complementary cumulative degree distribution for the amalgamation of the *Wooing of Étaín* and the *Dream of Óengus* fitted with a Weibull distribution (dashed red line) and a log normal (dotted blue line). Panel (b) shows the $\psi(\Delta)$ distribution. The disassortativity is clear in that characters show no preference to interact with characters of a similar degree.

in fig. 5.3 (b) where there is no clear peak for small difference in degree $\Delta$.

The amalgamated network is small world, but its transitivity, $C_T = 0.30$, is well predicted by $C_\text{n} = 0.21$, unlike what is found in the social networks in Chapter 3. The network robust to the random removal of nodes only. The community detection algorithm separates the two narratives and does not find any sub-communities with each.

### *5.1.4* Táin Bó Cúailnge

The *Táin* is the second largest of the Irish narratives with 422 characters. It also has the largest proportion of hostile edges among the Irish texts with more than 11% of the edges being hostile.

The degree distribution is shown in fig. 5.4 (a). Six characters have a degree above $k = 75$ and there are no characters in the range $35 < k < 75$. This makes it difficult to find a functional form for the degree distribution. The $\text{AIC}_\text{c}$ and BIC weights show the strongest support for a log-normal distribution but they also do not rule out a power law with exponent $\gamma = 2.27 \pm 0.08$ or a truncated power law with a slightly lower exponent of $\gamma = 2.16 \pm 0.04$ and $\kappa = 168.6 \pm 0.1$.

Due to the large disparity in degree, the network is disassortative with $r_k = -0.35 \pm 0.02$. This can be seen visually in fig. 5.4 (b) where we observe that the six most connected characters frequently interact with individuals with a degree difference of up to $\Delta = 150$. The assortativity for the friendly network is barely

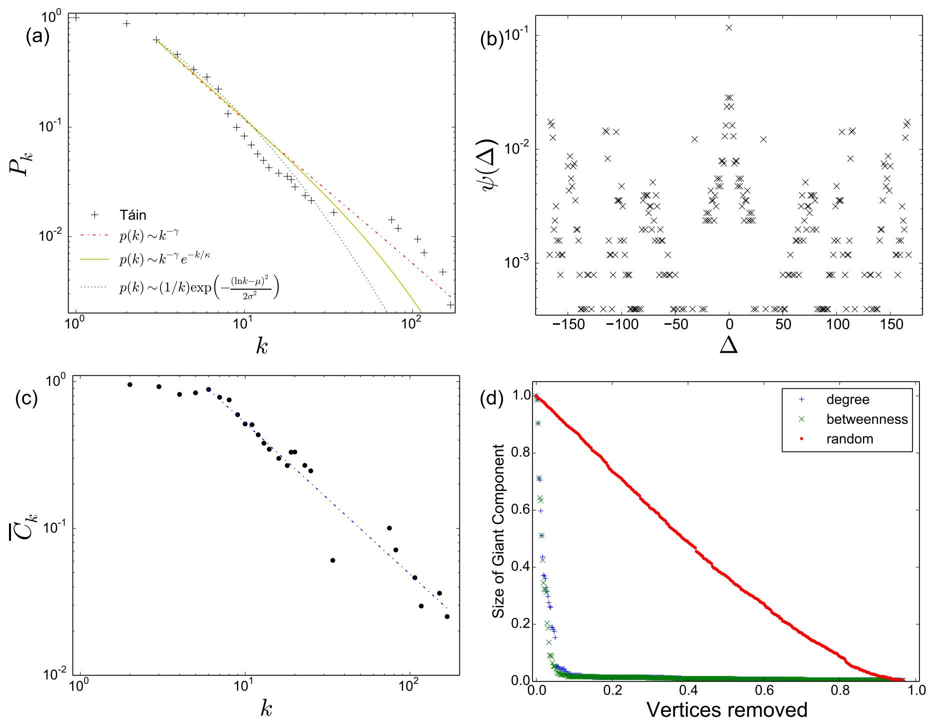



*Fig. 5.4:* Panel (a) shows the complementary cumulative degree distribution for the *Táin*. The most likely candidate model is a log normal distribution (dotted green line). The $AIC_c$ and BIC weights also give support to a power law (dash-dotted red line) and a truncated power law (continuous yellow line). Panel (b) shows the $\psi(\Delta)$ distribution. The high degree characters frequently interact with low degree characters resulting in multiple peaks with large $\Delta$. In panel (c) the mean clustering per degree is shown with a fitted power law. Panel (d) shows the relative size of the giant component as vertices are removed. It is robust to the random removal of vertices but fragile when removing by degree or betweenness.

affected with $r_k = -0.33 \pm 0.05$. The *Táin* is the only Irish myth whose network has a strong clustering disassortativity, $r_C = -0.30 \pm 0.03$. It has, however, a high Pearson similarity, $r_P = 0.14$. This is due to a large amount of characters sharing the same neighbours, namely the 6 highest degree vertices.

The Táin network is small world. It has a high clustering coefficient, however its transitivity, $C_T = 0.10$, is more than six times smaller than the naïve estimate, $C_{\rm n} = 0.62$. The mean clustering coefficient per degree $\bar{C}_k$ is shown in fig. 5.4 (c) on a log-log scale. The distribution is hierarchical and follows $\bar{C}_k \sim k^{-\beta}$, where $\beta = 1.03 \pm 0.06$.

The network is structurally balanced with less than 10% of the triangles con-

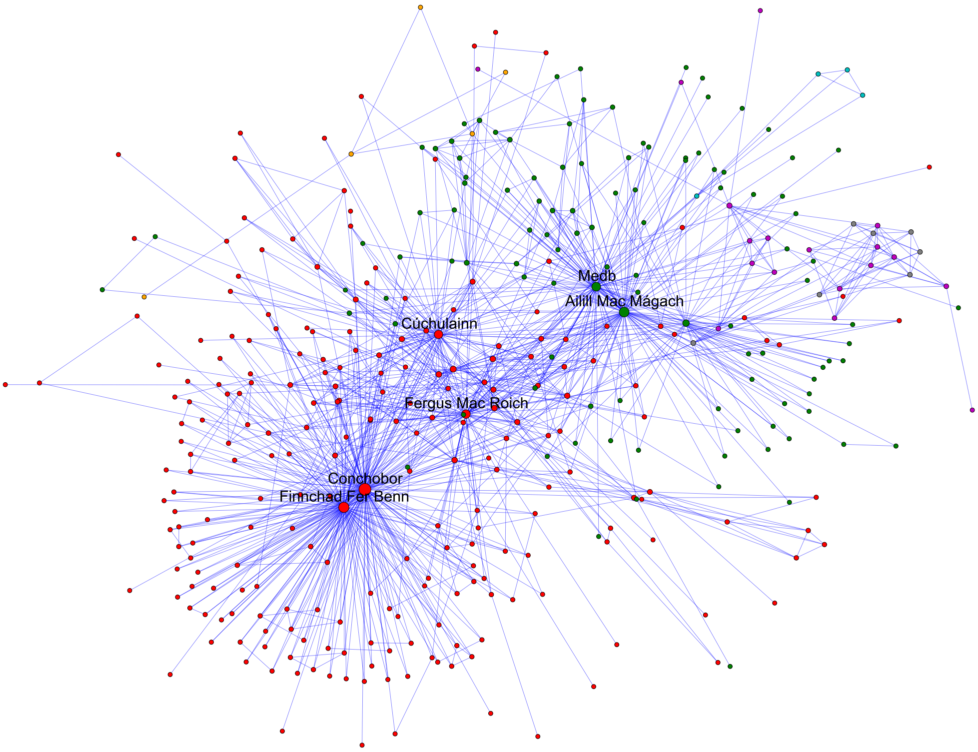



*Fig. 5.5:* The Girvan-Newman algorithm applied to the friendly network of the *Táin*. Different coloured vertices represent different communities. The two largest are the Ulster faction (red) and the Connacht faction (green).

taining an odd number of hostile edges. In fig. 5.4 (d), the robustness of the network is analysed. It is robust when vertices are randomly removed, however, the size of the giant component diminishes rapidly when vertices are removed by degree or betweenness.

The community detection algorithm is applied to the friendly network to give a better representation of the factions. The algorithm finds 6 communities with modularity $Q = 0.40$. These communities are shown with different colours in fig. 5.5. The two largest communities are the two opposing sides in the conflict. The 6 most connected characters are named in the figure and the algorithm assigns them all to the correct factions. Fergus Mac Roich however does change factions within the story and spends more of the story in the Connacht (green) faction than the Ulster (red) one.

It is clear from the degree distribution and the $\psi(\Delta)$ distribution that the 6 most connected characters have a large influence on the network. To investigate this

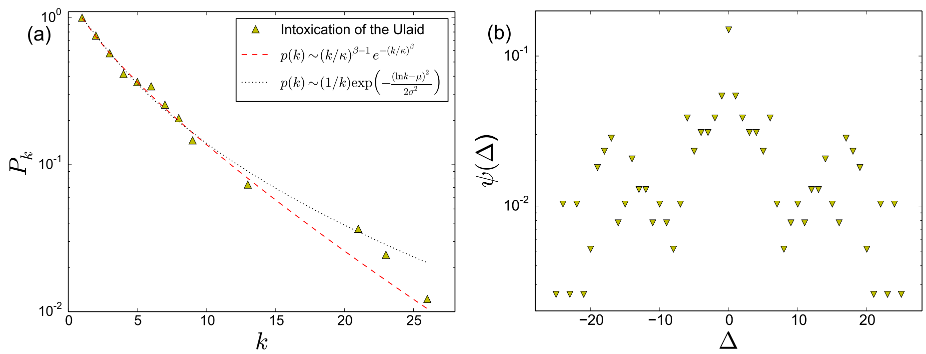



*Fig. 5.6:* Panel (a) shows the complementary cumulative degree distribution for the *Intoxication of the Ulaid*. The $AIC_c$ and BIC weights give almost equal support to a Weibull distribution (dashed red line) and a log-normal distribution (dotted black line). In panel (b) the $\psi(\Delta)$ distribution is shown. There are some fluctuations in at the extremities causing the network to become disassortative.

further we remove all edges with a weight of 1 associated with these six vertices. This leaves all the vertices intact but reduces the degree of the top 6 characters. The degree distribution decays faster with no support for a power law regime any longer. The most support is instead for a Weibull distribution.

The friendly network becomes neither assortative or disassortative with $r_k = -0.04 \pm 0.05$. The clustering assortativity also increases significantly with $r_C = 0.62 \pm 0.03$. The network remains small world with this change, however the transitivity, $C_T = 0.33$, is almost 9 times higher than naïvely estimated, $C_{\rm n} = 0.04$. Therefore by removing the weak interactions for the most connected characters, the network becomes more like the social networks of chapter 3.

### *5.1.5* Intoxication of the Ulaid

There are 82 characters in the network of the *Intoxication of the Ulaid*. Its degree distribution is shown in fig. 5.6 (a). It is well fitted by both a Weibull distribution and a log-normal distribution. The network is mildly disassortative with $r_k = -0.08 \pm 0.07$. This is visualised in fig. 5.6 (b).

It is a small world network and its transitivity, $C_T = 0.38$, is twice that of its naïve value, $C_{\rm n} = 0.19$. The network is structurally balanced and robust to the random removal of nodes only. It has an average clique size of 4.1 and the largest clique contains 7 vertices.

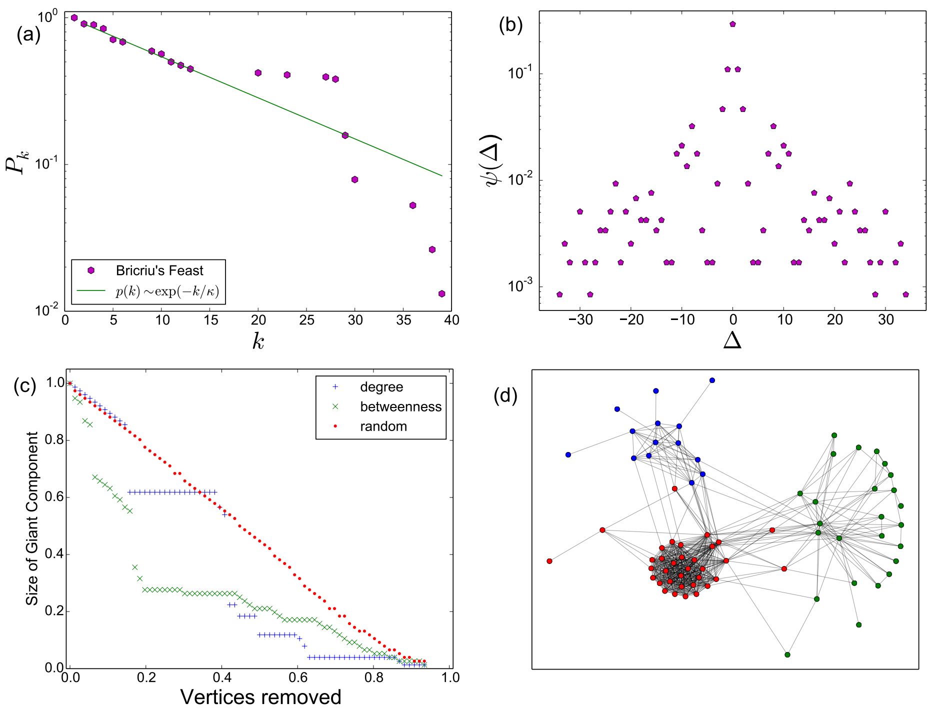



*Fig. 5.7:* Panel (a) shows the complementary cumulative degree distribution for *Bricriu's Feast* with an exponential distribution fitted as the continuous green line. In panel (b) the $\psi(\Delta)$ distribution is shown. This decays uniformly as the network is assortative. The robustness of the network is analysed in panel (c), the network is robust to the removal of vertices by degree until almost half of the vertices have been removed. Panel (d) shows three communities as identified by the Girvan-Newman algorithm with the clique (most of the red vertices) visible in the lower left.

### *5.1.6* Bricriu's Feast

*Bricriu's Feast* contains a clique of 29 Ulster warriors. These warriors all appear throughout many of the Ulster Cycle tales, however this is the only instance where they each interact with one another. This clique has a large effect on the properties of the network. For example, it has the largest average degree, $\langle k \rangle = 15.55$, of all the mythological networks studied here despite it having a comparatively low number of vertices, $N = 76$.

The degree distribution is shown in fig. 5.7 (a). It is not well fitted by any of the model distributions in section 2.2 but an exponential with parameter $\kappa = 15.5 \pm 0.2$ is displayed. The approximation $\kappa \approx \langle k \rangle$ when $k_{\text{min}} = 0$ from eq. (A.20) offers some support for an exponential distribution however.

The *Bricriu's Feast* network is highly assortative with $r_k = 0.48 \pm 0.05$, in contrast to the other Irish mythological networks studied here. The average degree of a vertex at the end of an edge from eq. (2.60) is $\mathrm{E}[\mathrm{k}] = 24.90$. This is less than the degree of each character in the clique. Therefore when calculating the assortativity, the degree at either end of all edges in the clique is above $\mathrm{E}[\mathrm{k}]$ which drives the assortativity. Fig. 5.7 (b) shows the $\psi(\Delta)$ distribution in which there are very few fluctuations here when compared to what is observed in disassortative networks. This network also has a high clustering assortativity, $r_C = 0.40 \pm 0.04$, and a Pearson similarity of $r_P = 0.10$.

The network is small world. It has a high clustering coefficient, $C = 0.70$, and transitivity, $C_T = 0.84$, highlighting an abundance of triangles, most of which are contained in the clique. The transitivity however, is not much larger than naïvely estimated, in contrast to the social networks in chapter 3. The clique also makes the network very robust to targeted removal of nodes by degree as shown in fig. 5.7 (c). Randomly removing 25% of the vertices (19 vertices) leaves the giant component 94.7% intact, removing the 19 characters with the highest degree leaves the giant component 82.5% intact but removing them by highest betweenness leaves it at just 36.8%.

The Girvan-Newman algorithm detects three communities with a modularity of $Q = 0.32$. These are displayed in fig. 5.7 (d) where the Ulster warriors are coloured red. The other two communities correspond to some Ulster characters in other households and the characters who judge the three Ulster heroes.

### 5.1.7 Amalgamated Selection of Ulster Cycle Texts

Many of the Ulster cycle narratives contain too few characters for reliable network analysis but they contain recurring characters. Therefore the texts where characters are named consistently are merged to form a larger Ulster Cycle network. The texts included in this network are as follows; the *Cattle Raid of Fróech*, the *Wasting Sickness of Cú Chulaind*, the *Tale of Macc Da Thó's Pig*, the *Intoxication of the Ulaid*, *Bricriu's Feast*, and the *Táin* pre-tales (translated by a different author); the *Exile of the Sons of Uisliu*, the *Boyhood Deeds of Cú Chulaind*, the *Twins of Macha*, the *Death of Aife's only son* and the *Birth of Cú Chulaind*. The *Táin* itself is not included as characters' names are spelled differently to other translations. There are too many characters to reliably identify each correctly here.

This amalgamated Ulster Cycle network contains 229 unique characters with an average degree of $\langle k \rangle = 9.34$. The degree distribution is shown in fig. 5.8 (a).

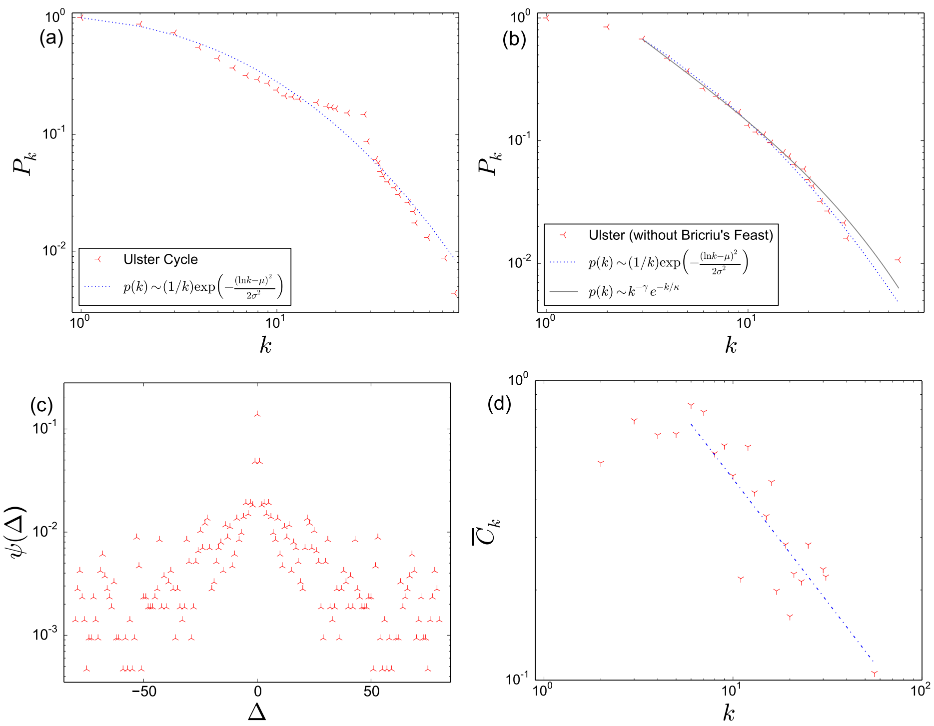



*Fig. 5.8:* Panel (a) shows the complementary cumulative degree distribution on a log-log scale for the Ulster Cycle with a log-normal distribution fitted as the dotted blue line. In panel (b) *Bricriu's Feast* has been removed from the network. The lack of the large clique smoothens the distribution and support is given for truncated power law (continuous grey line) and a log-normal distribution (dotted blue line). The $\psi(\Delta)$ distribution is shown in panel (c). This is neither assortative or disassortative. Panel (d) shows the mean clustering per degree $\bar{C}_k$ with a fitted power law for the network without Bricriu's feast. Its decay indicates the network has a hierarchical structure.

It is best fitted by a log-normal distribution with parameters $\mu = 1.53 \pm 0.06$ and $\sigma = 1.23 \pm 0.05$. The clique from *Bricriu's Feast* has an affect on the network which can clearly be seen in the degree distribution at $k = 28$. Upon removing *Bricriu's Feast* from the Ulster Cycle network, the distribution becomes smoother as shown in fig. 5.8 (b). The AIC$_c$ and BIC weights offer strongest support for a truncated power law with parameters $\gamma = 1.90 \pm 0.07$ and $\kappa = 56.59 \pm 0.01$. A log-normal distribution is still not ruled out however.

The removal of *Bricriu's Feast* from the network leaves 187 unique characters with an average degree of $\langle k \rangle = 5.70$. The loss of the clique also reduces the assortativity from $r_k = -0.03 \pm 0.04$ to $r_k = -0.13 \pm 0.03$. As with the *Bricriu's*

*Feast* network the clique has a large effect on the assortativity of the amalgamation. The $\psi(\Delta)$ distribution for the entire network is shown in fig. 5.8 (c). This network has no significant correlations between the degrees of interacting vertices, therefore we do not observe large fluctuations in the tails of the distribution like the disassortative networks, nor do we observe the uniform decay about the mean, as with assortative networks.

The network is small world (with or without *Bricriu's Feast*) and shows evidence of hierarchical structure. This is more clear in the case without *Bricriu's Feast* as shown in fig. 5.8 (d). Here, a power law with exponent $\beta = 0.82 \pm 0.11$ is fitted to the distribution. The network is also structurally balanced with 9% of the triangles containing an odd number of hostile edges.

The Girvan-Newman community detection algorithm does not have much success finding communities. The modularity peaks at $Q = 0.23$ with 15 communities. Removing *Bricriu's Feast* however yields a modularity of $Q = 0.34$ for 7 communities and a higher $Q = 0.44$ for 13 communities. There are 9 distinct narratives without *Bricriu's Feast*, however the algorithm does not successfully split them apart due to the abundance of overlapping characters.

Of the ten Ulster Cycle narratives, *Bricriu's Feast* is the most distinct of the networks and has a large effect on the amalgamated network. However, it does not seem pertinent to leave it out. In each of the Ulster Cycle narratives we are introduced to various Ulster heroes, many of whom appear at one point or another in the *Táin* too (which is not in the amalgamated network). It does seem likely that all these characters from the same time who serve the same king interact with each other. However only *Bricriu's Feast* makes that clear.

It is also interesting to note that with *Bricriu's Feast*, the network looks more like the social networks studied in chapter 3; it is not disassortative, a log-normal is the most likely degree distribution and it has an average path length slightly longer than that of its random counterpart, unlike that of most of the other Irish myths. This implies that the social networks in most of these myths do not give a good representation of the societies they portray and only one of these narratives catches a glimpse of this society. As each story is just a snapshot of characters' interactions, this network captures more about the type of narrative rather than the social network at the time. If this is indeed the case, this has strong implications for what information the networks in this study portray. This point will be returned to in section 6.2.

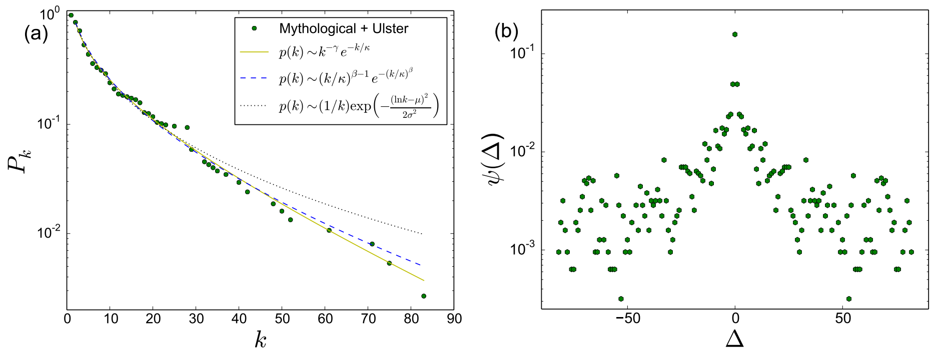



*Fig. 5.9:* Panel (a) shows the complementary cumulative degree distribution for the amalgamation of the Mythological and Ulster Cycles translated in Gantz (1981). The $AIC_c$ and BIC weights give almost equal support to a truncated power law (continuous yellow line), a Weibull distribution (dashed blue line) and a log-normal distribution (dotted black line). In panel (b) the $\psi(\Delta)$ distribution is shown, this network is also assortative or disassortative.

### *5.1.8 The Mythological and the Ulster Cycle*

As mentioned in section 4.3.1, the Mytholocial and Ulster Cylces in Irish mythology contain many overlapping characters. The following network merges all 13 narratives from Gantz (1981). These are the three Mythological Cycle narratives above: the *Wooing of Étaín*, the *Destruction of Da Derga's Hostel* and the *Dream of Óengus*, and the 10 Ulster Cycles tales in the previous section.

The combined network contains 374 characters with 1574 interactions. Of these, only 52 edges are hostile. The clique from *Bricriu's Feast* also has an impact on this network as shown at $k = 28$ on the degree distribution in fig. 5.9 (a). The $AIC_c$ and BIC weights give support for a truncated power law, a Weibull distribution and a log-normal distribution. However if *Bricriu's Feast* is removed, a log-normal distribution with parameters $\mu = 1.15 \pm 0.05$ and $\sigma = 1.11 \pm 0.03$ becomes the most likely of the candidate models.

The assortativity of the network is $r_k = 0.01 \pm 0.03$, again indicating that there are no correlations between neighbours' degrees. The $\psi(\Delta)$ distribution is shown in fig. 5.9 (b) highlighting this lack of correlation. However, as with the previous network, when *Bricriu's Feast* is removed, the assortativity drops to $r_k = -0.14 \pm 0.03$. The clustering assortativity $r_C = 0.21 \pm 0.03$ and Pearson similarity $r_P = 0.02$ are unaffected by the removal of *Bricriu's Feast*.

In either case, the network is small world and is fragile to the removal of nodes

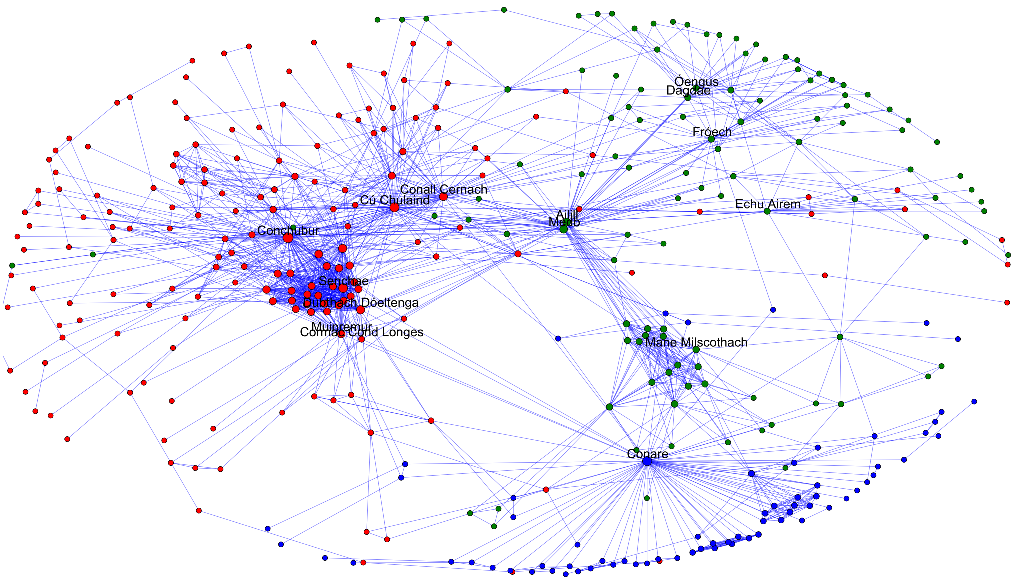



*Fig. 5.10:* The Girvan-Newman algorithm applied to the amalgamated network of the Mythological and Ulster Cycles. The red vertices are characters from the Ulster faction and the green vertices are characters from the Tuatha Dé Danann in the Mythological cycle and the Connacht faction. The blue nodes are characters who appear in the text the *Destruction of Da Derga's Hostel*.

by degree or betweenness. Applying the community detection algorithm, 3 communities are found with modularity $Q = 0.50$. These communities are shown in fig. 5.10. The vertices coloured in red are mostly Ulster warriors. The green vertices are a mixture of characters from Connacht and the mythological characters from the Tuatha Dé Danann. Finally the blue vertices are mostly characters from the tale the *Destruction of Da Derga's Hostel*. In spite of this network containing 13 tales, the Ulster and Connacht factions are still easily separable, however most of the different narratives are not.

One final observation is that the network properties of the amalgamation of Irish myths does not drastically change as additional texts are added. The properties of most of these myths alone are not too dissimilar from one another with the exception of *Bricriu's Feast*. The tales are mostly centred on a small number of protagonists and usually involve them going on a journey.

### 5.1.9 *The* Colloquy of the Ancients

The network of the *Colloquy of the Ancients* is the largest of the myths with 732 characters. The average degree of the network is $\langle k \rangle = 6.20$. The degree distribution for the network is shown in fig. 5.11 (a). It is fitted with both a truncated power law (with parameters $\gamma = 1.61 \pm 0.03$ and $\kappa = 79.89 \pm 0.01$) and a log-normal (with parameters $\mu = 0.00 \pm 0.05$ and $\sigma = 1.67 \pm 0.06$). The two highest degree vertices are Fionn mac Cumhaill ($k = 127$) and Cáilte ($k = 102$). Though these two high degree vertices frequently interact with characters with much lower degrees, the network also contains a clique of 41 vertices. These two factors balance each other out, causing the network to have no degree-degree correlations, $r_k = -0.01 \pm 0.02$. There are peaks for large $\Delta$, but even larger peak about $\Delta = 0$ as displayed in fig. 5.11 (b). The network is also clustering assortative ($r_C = 0.48 \pm 0.02$) and has a low Pearson similarity measure ($r_P = 0.02$).

The average path length for the *Colloquy of the Ancients*, $\ell = 3.66$, is close to its random counterpart, $\ell_{\text{rand}} = 3.80$ and its clustering coefficient, $C = 0.62$ is much greater, $C_{\text{rand}} = 0.01$. This is the small world property. The transitivity, $C_T = 0.51$, is reasonably well estimated by the naïve prediction, $C_{\text{n}} = 0.23$. Due to the large clique, the network is more robust to the removal of vertices by degree than normal however it is still fragile when nodes are removed in order of betweenness. Removing 10% of the vertices (73 characters) by betweenness the giant component is 3% of its original size. However removing by degree it is 31%, and randomly 72%, of its original size.

The network of the *Colloquy of the Ancients* has similar properties to the amalgamated selection of Ulster Cycle narratives. They each contain a large clique, but also interacting characters with large difference in degrees resulting in an absence of assortativity. Their degree distributions are both well fitted by log-normal distributions.

### 5.1.10 Fianaigecht

*Fianaigecht* contains further tales of the Fianna. It contains 193 vertices with an average degree of $\langle k \rangle = 3.50$. This network contains a genealogy of 55 characters going from a king of Ireland back to God, each with a degree of 2. Removing this, the size changes to $N = 138$ and the average degree increases to $\langle k \rangle = 4.10$. This is referred to as "*Fianaigecht* (no gene)" in table 5.1 and table 5.2.

The degree distribution with the genealogy is shown in fig. 5.12 (a). The most

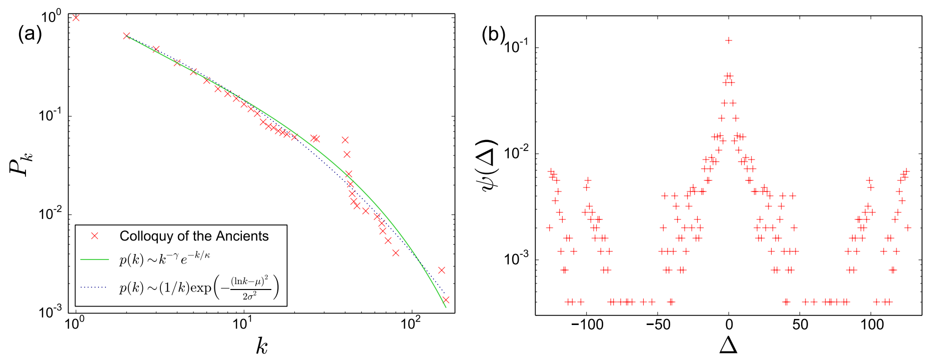



*Fig. 5.11:* Panel (a) shows the complementary cumulative degree distribution for the *Colloquy of the Ancients*. The most likely candidate models are a truncated power law (green continuous line) and a log normal distribution (dotted blue line). Two characters have higher degrees than estimated from the fits. Panel (b) shows the $\psi(\Delta)$ distribution. The two highest degree characters frequently interact with low degree characters resulting in peaks with large $\Delta$. This causes the disassortativity in the network.

likely of the candidate models is a truncated power law with parameters $\gamma = 2.21 \pm 0.06$ and $\kappa = 53.59 \pm 0.01$. There is also strong support for a power law with exponent $\gamma = 2.40 \pm 0.12$. Without the genealogy, a truncated power law is still supported but a Weibull distribution and a log-normal distribution are also likely.

The protagonist of this text is again Fionn Mac Cumhaill who has a degree of $k = 53$. The second highest character (Cáilte again) has a degree of $k = 19$. This large difference in degree has implications for the assortativity, $r_k = -0.07 \pm 0.03$. Fionn's interaction with low degree characters are clearly represented in fig. 5.12 (b). The genealogy also affects the assortativity as each character in it has the same degree. Removing the genealogy, the network becomes more disassortative with $r_k = -0.14 \pm 0.04$. The network has a high clustering assortativity in each case but when removing the genealogy, the Peason similarity increases from its original value of $r_P = 0.03$ to $r_p = 0.05$, as vertices share neighbours more commonly.

The genealogy however has its largest effect on the network's average path length, $\ell = 17.48$, and diameter, $\ell_{\max} = 64$. Removing the genealogy reduces these to $\ell = 3.41$ and $\ell_{\max} = 10$. The clustering coefficient also goes from $C = 0.29$ to $C = 0.41$. Without the genealogy the network is small world. The clustering transitivity is close to its naïve estimate in each case. Fig. 5.12 (c) shows the decay of the clustering coefficient with degree. It is fitted with a power law

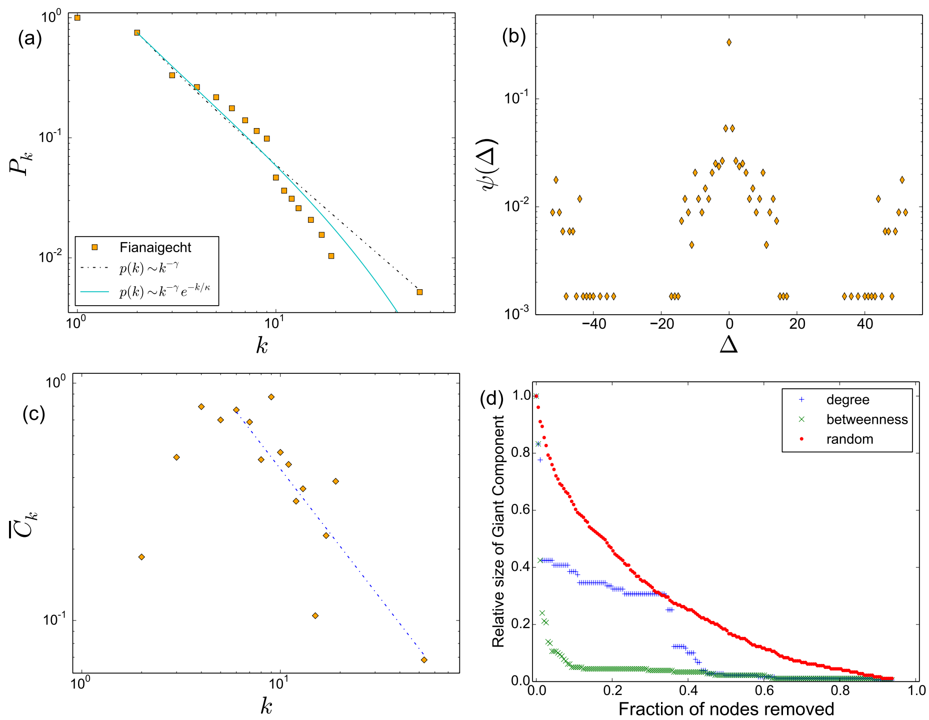



*Fig. 5.12:* Panel (a) shows the complementary cumulative degree distribution for the *Fianaigecht*. The most likely candidate models are a power law (black dash-dotted line) or a truncated power law (continuous cyan line). Panel (b) shows the $\psi(\Delta)$ distribution. The protagonist frequently interacts with low degree characters as seen at the ends of the plot. In panel (c) the mean clustering per degree is shown with a fitted power law. Panel (d) shows the relative size of the giant component as vertices are removed. After the loss of the three most connected characters, it becomes robust to the removal of vertices by degree.

with exponent $\beta = -1.08 \pm 0.20$.

Upon removing the three most connected vertices, the network becomes very robust to the removal of vertices by degree, as shown in fig. 5.12 (d). This is because the text contains 6 different tales which are rather loosely connected. These three Fianna warriors tie the different narratives together. This is investigated in more detail by using a community detection algorithm. One of the 6 tales in *Fianaigecht* only contains two characters. With the exception of this, the community detection algorithm succeeds in separating the 5 components with $Q = 0.61$.

Like the *Colloquy of the Ancients*, the *Fianaigecht* network relies heavily on its protagonist. For example removing Fionn the assortativity jumps to $r_k \pm 0.32 \pm 0.03$.

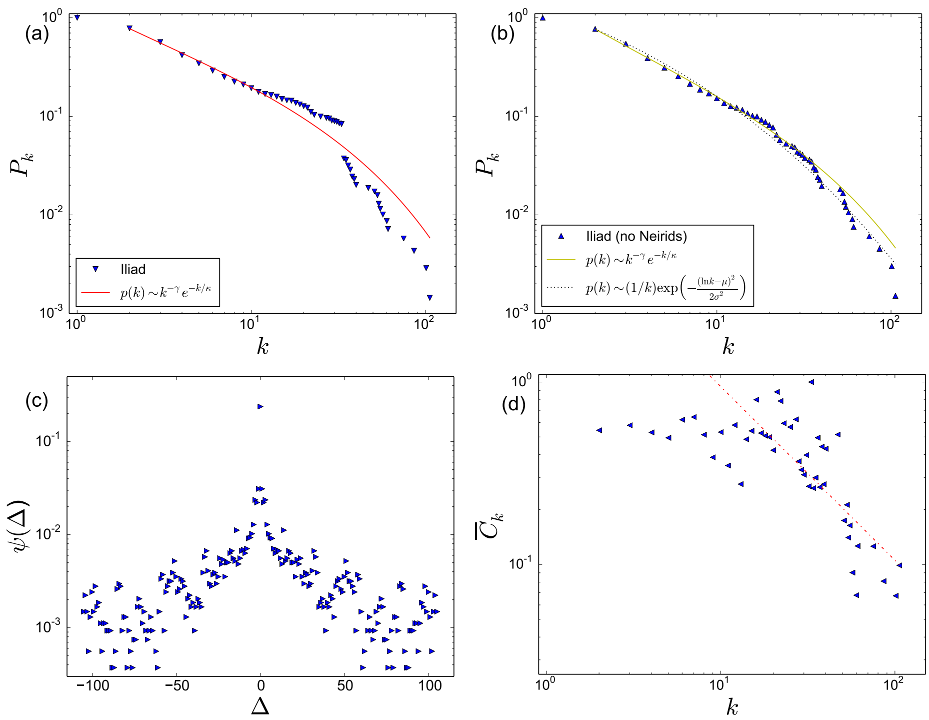



*Fig. 5.13:* Panel (a) shows the complementary cumulative degree distribution for the *Iliad*. The most likely candidate model is a truncated power law (continuous red line). After removing the clique at $k = 33$, the degree distribution is shown again in panel (b). The most likely model again is a truncated power law (continuous yellow line) however mild support is offered for a log-normal distribution (dotted black line). Panel (c) shows the $\psi(\Delta)$ distribution. It is only mildly disassortative due to the long tails. In panel (d) the mean clustering per degree is shown with a fitted power law.

## 5.2 *Greek & Roman Mythology*

The networks of classical mythology have similar properties. They are all small world and disassortative. However, we see below that there are differences due to the number of protagonists in the *Iliad* which affects the degree distribution.

### *5.2.1* Iliad

The *Iliad* has the largest number of edges of all the single mythological networks constructed here. It contains 694 vertices with 2684 edges. Of these, 355 interactions are purely hostile. The network contains a clique of 34 Neirids (sea-nymphs). These all appear to comfort Achilles' mother Thetis. When showing the degree distribution (fig. 5.13 (a)), these cause a large jump at $k = 33$. The most likely fit for the degree distribution of the candidate models is a truncated power law with parameters $\gamma = 1.52 \pm 0.03$ and $\kappa = 82.59 \pm 0.01$. The Neirids who are only linked to other Neirids are then removed and shown in fig. 5.13 (b). A truncated power law is still the most likely distribution with a parameters $\gamma = 1.69 \pm 0.03$ and $\kappa = 105.2 \pm 0.1$. The $\text{AIC}_\text{c}$ and BIC weights also offer mild support for a log-normal.

The network is mildly disassortative, $r_k = -0.08 \pm 0.02$. Fig. 5.13 (c) shows some fluctuations in the tail of the $\psi$-distribution. The heroes in the *Iliad* tend to have very large hostile degrees (which are essentially kill counts). As this network has a relatively large number of hostile links we examine the assortativity for just the friendly edges. This yields $r_k = 0.09 \pm 0.02$ making it assortative. In contrast the hostile network is disassortative with $r_k = -0.45 \pm 0.05$. Removing the Neirids reduces the friendly network's assortativity slightly to $r_k = -0.01 \pm 0.02$ but it is still neither assortative for disassortative. The *Iliad* is clustering assortative, $r_C = 0.53 \pm 0.02$, but has a low Pearson similarity measure, $r_P = 0.02$.

The *Iliad* network is small world and has slightly higher transitivity than naïvely expected. The mean clustering coefficient per degree is shown in fig. 5.13 (d). Its decay shows evidence of hierarchical structure in the network. It is fitted with a power law with exponent $\beta = 0.95 \pm 0.13$. The network is also structurally balanced with less than 2% of its triangles containing an odd number of hostile edges.

The Girvan-Newman community detection algorithm finds 3 communities with $Q = 0.60$ for the friendly network of the *Iliad*. These three communities (shown in fig. 5.14) correspond to the two opposing factions, the Greeks and the Trojans, and the third group is the Neirids. The modularity reaches a plateau $Q = 0.66$ with 12 communities. Eight of these communities have 22 or fewer vertices. The four larger ones are the Neirids, the gods, the Greeks and the Trojans.

### *5.2.2* Odyssey

The network of the *Odyssey* contains 301 vertices ad just 30 of the 1019 edges are purely hostile. The protagonist Odysseus has a degree of $k_{\max} - 112$ which much larger than the next highest degree, that of his son Telemachus $k = 43$. This large jump is clearly illustrated in the degree distribution in fig. 5.15 (a). A Weibull distribution with parameters $\beta = 0.70 \pm 0.04$ and $\kappa = 4.27 \pm 0.02$ receives the strongest support. However there is also moderate support for an exponential when Odysseus is excluded. An exponential is shown when the highest two degree

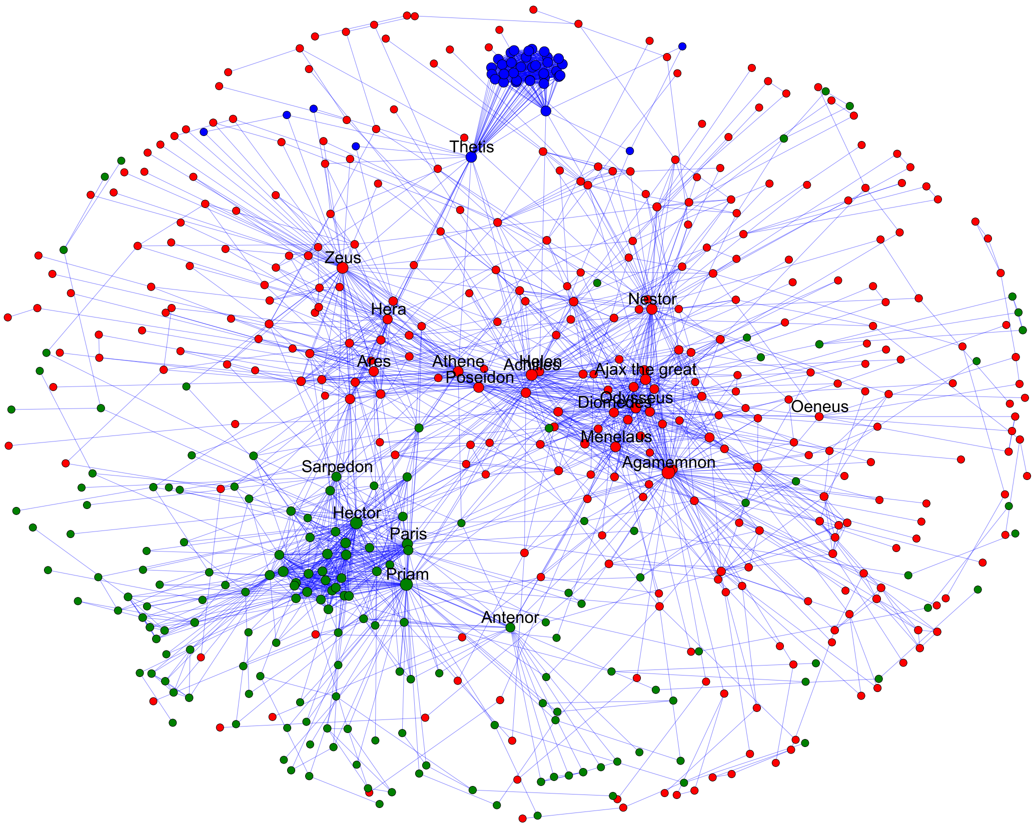



*Fig. 5.14:* The Girvan-Newman algorithm finds 3 communities in the friendly network of the *Iliad*. The red vertices are from the Greek faction, the green vertices are Trojan characters. The blue nodes are Neirids.

characters are excluded, this has an exponent of $\kappa = 5.78 \pm 0.01$.

The network is disassortative with $r_k = -0.08 \pm 0.02$. Fig. 5.15 (b) displays the $\psi(\Delta)$ distribution where we observe Odysseus' frequent interactions with minor characters on his journey. The network is clustering assortative, $r_C = 0.38 \pm 0.03$, which is in part due to to an abundance of cliques. The network has 171 cliques, the largest of which contains 17 vertices. The distribution of cliques $P_{C_s}$ is depicted in fig. 5.15 (c) with a fitted exponential showing its fast decay.

The network is small world with $\ell = 3.29$ and $C = 0.45$, similar values to those of the *Iliad*. The transitivity, $C_T = 0.38$, is a little larger than the naïve estimation, $C_{\text{n}} = 0.15$. The network is surprisingly robust to the removal of vertices by degree as shown in fig. 5.15 (d). Removing the 30 highest degree vertices, the giant component remains at 76% of its original size, however in the case of

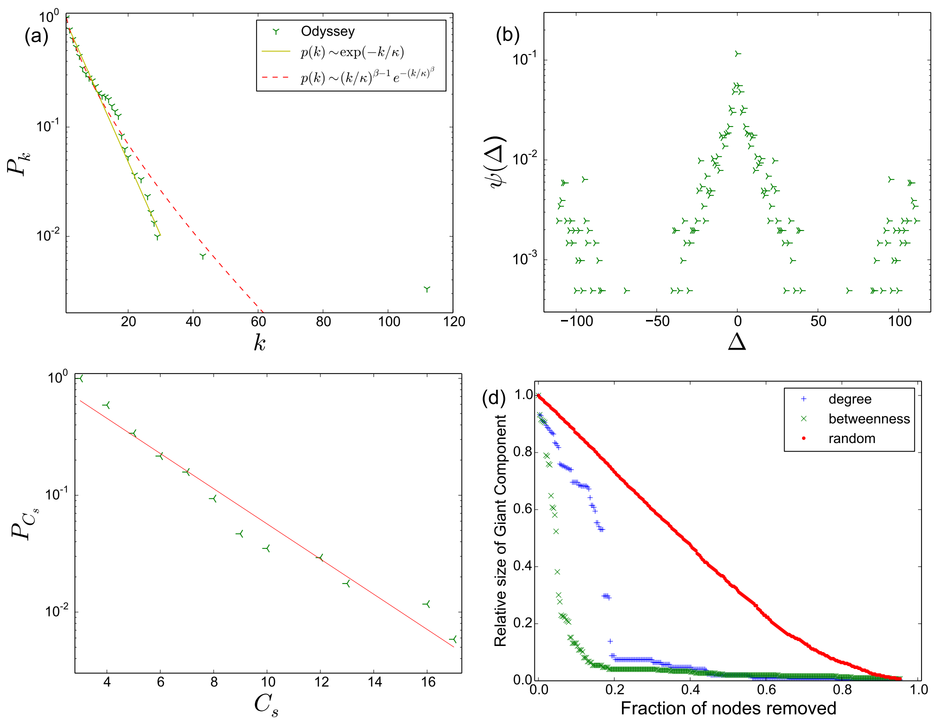



*Fig. 5.15:* Panel (a) shows the complementary cumulative degree distribution for the network of *Odyssey*. The most likely candidate model is a Weibull distribution (dashed red line). There is moderate support for an exponential also (continuous yellow line) which does not fit to the two highest degree vertices. Panel (b) shows the $\psi(\Delta)$ distribution. Odysseus, the protagonist, frequently interacts with low degree characters as seen at the ends of the plot. The clique size distribution $P_{C_s}$ is shown with a fitted exponential in panel (c). Panel (d) shows the robustness of the *Odyssey* network. It is initially robust to the removal of high degree vertices but very fragile when nodes are removed by betweenness.

betweenness this is just 14%.

With the exception of the degree distribution the *Odyssey* has many similar network properties to those of the *Iliad*. A core difference is that the *Odyssey* has single protagonist rather than multiple which affects the degree distribution. Excluding Odysseus from the degree distribution, we observe that an exponential fits the data well, this combined with the robustness upon degree removal shows that the network is not reliant on its protagonist unlike the case of networks with power law distributions.

A final note on the *Odyssey* is that Miranda *et al.* (2013) also analysed the social network of this which yields different results to our own. A difference between

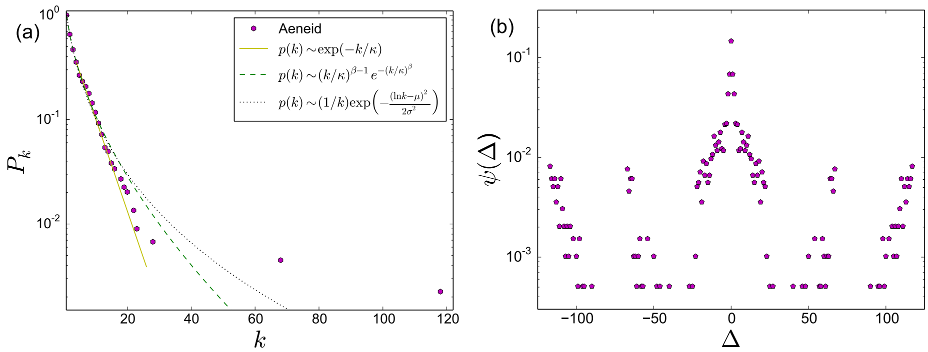



*Fig. 5.16:* Panel (a) shows the complementary cumulative degree distribution for the network of *Aeneid*. The most likely candidate models are a Weibull distribution (dashed green line) and a log-normal distribution (dotted black line). There is moderate support for an exponential also (continuous yellow line) which does not fit to the two highest degree vertices. Panel (b) shows the $\psi(\Delta)$ distribution for the disassortative network.

the number of vertices is most likely due to the fact that we exclude characters with degree $k = 0$ from the network. Therefore when we report the number of vertices, it is not necessarily the same as the number of names in the text. A second key difference is in how the edges are created. Miranda *et al.* (2013) create an edge if one characters cites another. We only do this when it is clear both characters have met. This leads to a large difference in the number of edges. For example characters regularly say Zeus' name, we do not create an edge between the character and Zeus unless it is explicit that they have interacted. Similarly, when characters die they are often said to go to "the House of Hades". At one point in the tale, Odysseus even travels to the Underworld and meets many of the dead. We do not create a link between a deceased character and Hades, therefore Hades and Zeus have relatively low degrees in our network. As shown by Miranda *et al.* (2013) however, Zeus is the second most connected characters, and Hades has the seventh highest degree in their network. This accounts for the large deviation in the network properties.

### 5.2.3 Aeneid

There 444 unique characters in the *Aeneid* with an average degree of $\langle k \rangle = 4.42$. As in the *Odyssey* previously however, two of these characters have particularly high degrees in comparison to the others. These are Aeneas $k = 118$ and Turnus

$k = 68$. The degree distribution in fig. 5.16 (a) highlights this. The $\text{AIC}_\text{c}$ and BIC weights give the almost equal support for a Weibull distribution ($\beta = 0.55 \pm 0.02$ and $\kappa = 1.48 \pm 0.03$) and a log-normal distribution ($\mu = 0.62 \pm 0.04$ and $\sigma = 1.21 \pm 0.04$). However, again like the *Odyssey*, an exponential ($\kappa = 4.87 \pm 0.02$) receives moderate support when fitting to all but the two highest degree vertices.

The network is disassortative with $r_k = -0.13 \pm 0.02$). This is clearly shown by the large peaks for difference of degree $|\Delta| > 50$ in the $\psi(\Delta)$ in fig. 5.16 (b). It is clustering assortative $r_C = 0.33 \pm 0.03$. The *Aeneid* is small world with $\ell = 3.57$ and $C = 0.41$. However its transitivity, $C_T = 0.21$, is close the naïve estimate, $C_\text{n} = 0.13$.

## 5.3 Germanic Mythology

The Germanic myths contain different types of networks. The heroic epics like *Beowulf*, *Völsungasaga* and the *Nibelungenlied* are small world and disassortative. The *Edda*s and the *Orkneyinga saga* contain many different narratives giving them longer than expected average path lengths. Specific details of each are given below.

### 5.3.1 Beowulf

The network of *Beowulf* contains 72 characters and 167 edges. Of these, 26 edges are hostile. Two of the characters have disproportionately large degrees as shown in fig. 5.17 (a). These are the eponymous protagonist Beowulf and Hrothgar, the king of the Danes. The $\text{AIC}_\text{c}$ and BIC weights for the candidate models provide strongest support for a log-normal distribution with parameters $\mu = 1.26 \pm 0.07$ and $\sigma = 0.77 \pm 0.04$. There is also mild support for an exponential distribution with $\kappa = 4.12 \pm 0.04$.

The network is disassortative with $r_k = -0.12 \pm 0.06$. As shown in fig. 5.17 (b), the two highest degree vertices frequently interact with low degree characters. The *Beowulf* network has low clustering assortativity with $r_C = -0.05 \pm 0.09$ and a low Pearson similarity measure.

*Beowulf* is small world and structurally balanced with just 4% of the triangles containing an odd number of hostile edges. The mean clustering coefficient per degree is displayed in fig. 5.17 (c). It is fitted with a power law with exponent $\beta = 0.83 \pm 0.11$. The network component is less than 70% connected however. This is due to two tales in the story dealing with events from the past. It is fragile to the removal of vertices in order highest degree and betweenness (see fig. 5.17 (d)).

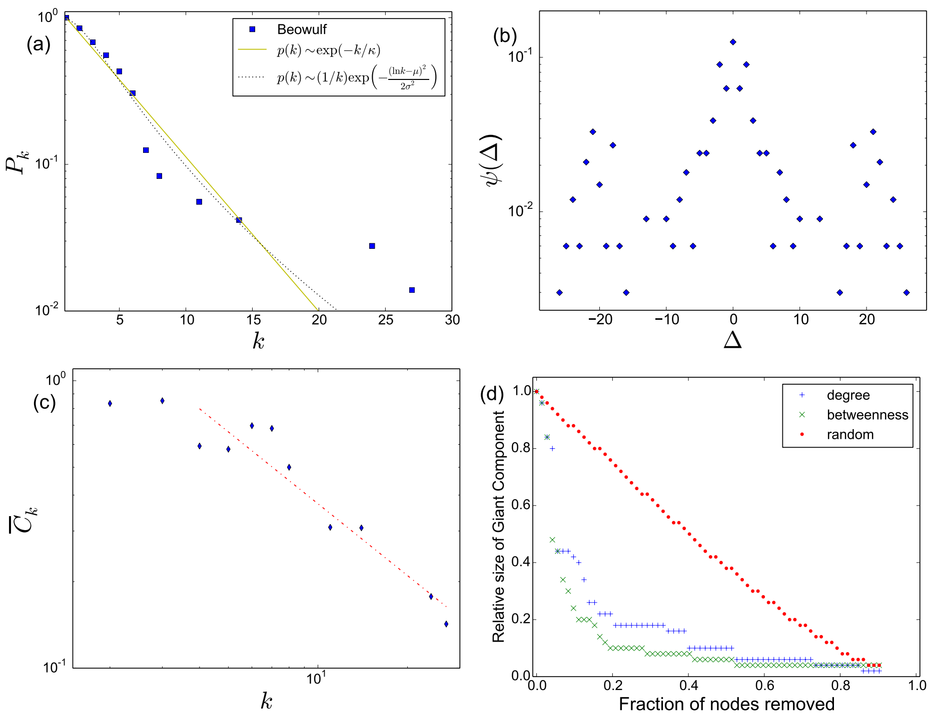



*Fig. 5.17:* Panel (a) shows the complementary cumulative degree distribution for the network of *Beowulf*. The most likely candidate model is a log normal distribution (dotted black line). There is moderate support for an exponential also (continuous yellow line). Panel (b) shows the $\psi(\Delta)$ distribution. The protagonist frequently interacts with low degree characters as seen at the ends of the plot. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the robustness of the *Beowulf* network. It is very fragile when vertices are removed by degree or betweenness.

The community detection algorithm finds 5 components (with $Q = 0.40$) in friendly network as displayed in fig. 5.18. In the figure the red vertices are Geats and the green nodes are the Danes that Beowulf goes to aid. The two orange vertices are ancestors of Beowulf. The blue vertices are the Swedes who had been at war with the Geats and the grey nodes are the characters who were involved in the incident with the dragon towards the end of the narrative.

### 5.3.2 Poetic Edda

The *Poetic Edda* contains three large cliques that dominate the properties of the network. The god Heimdall creates the structure of human society by creating three different houses, the lowest being the house of a Thrall, the next a Farmer

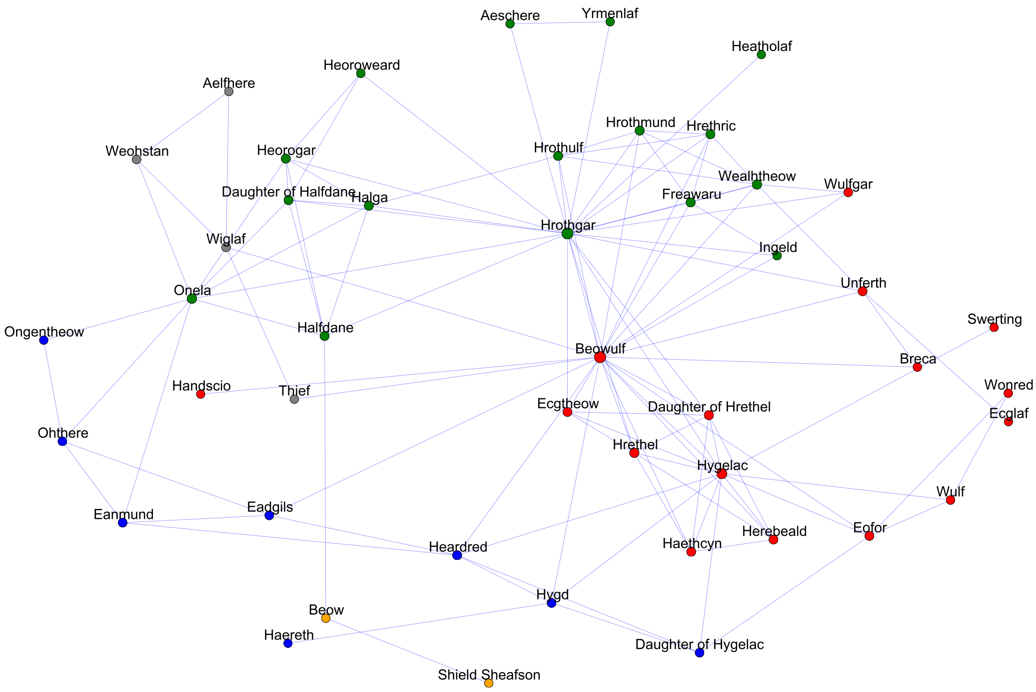



*Fig. 5.18:* The Girvan-Newman algorithm finds 5 communities in the friendly giant component of *Beowulf*.

and the highest a Lord. Each house has a large family to continue their line. These cliques for each house contain 19, 24 and 19 vertices respectively. There is a fourth large clique containing 17 characters. The complete network contains 354 characters and has an average degree of $\langle k \rangle = 6.43$.

The degree distribution is displayed in fig. 5.19 (a). A log-normal and Weibull distribution receive almost equal support from the $\text{AIC}_\text{c}$ and BIC weight but there is also moderate support for a truncated power law. None of these fit the data particularly well however. This is due to the cliques causing a large jump in the tail of the degree distribution.

The network is very strongly assortative with $r_k = 0.70 \pm 0.03$. Again, this is a result of the cliques. Almost every character in the clique interacts with nobody outside of that clique. Therefore the degree of each vertex is on the same side of the average degree at the of an edge $\text{E}[k]$ thus positively contributing to the assortativity. The corresponding $\psi(\Delta)$ distribution is shown in fig. 5.19 (b). The network is also strongly clustering assortative $r_C = 0.64 \pm 0.03$. Again each member of a clique shares the same clustering coefficient.

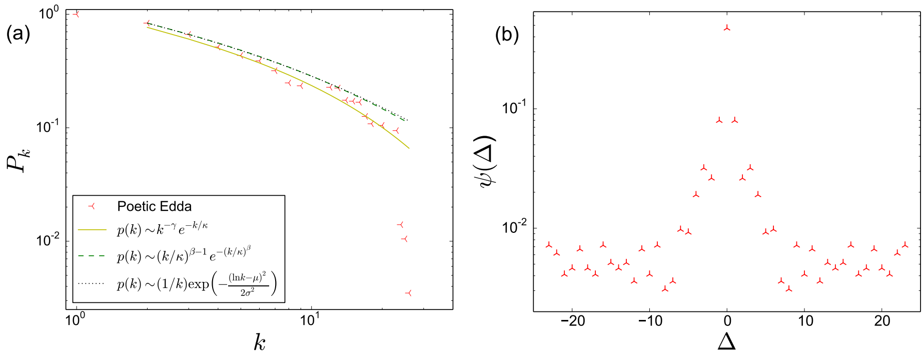



*Fig. 5.19:* Panel (a) shows the complementary cumulative degree distribution for the *Prose Edda* network. The most likely candidate model is a log-normal distribution (dotted black line) and Weibull distribution (dashed green line). There is moderate support for a truncated power law also (continuous yellow line). Panel (b) shows the $\psi(\Delta)$ distribution. The protagonist frequently interacts with low degree characters as seen at the ends of the plot.

The average path length of the *Poetic Edda* is $\ell = 4.99$ which is large compared to average path length of its random graph $\ell_{\text{rand}} = 3.34$. Therefore the network is not small world. The *Poetic Edda* comprises many different texts that are only loosely connected. This causes the large value for the average path length and also a strong community structure. The transitivity is also particularly high, $C_T = 0.84$, compared to the naïve prediction of $C_{\text{n}} = 0.08$. Once again this mostly due to an abundance of triangles in the cliques.

The network is not well connected as shown from the giant component containing 69.2% of the vertices. Removing 18 of the 354 (5%) vertices leaves the largest component containing just 24 vertices. As a result the community detection algorithm finds 10 communities with a high modularity of $Q = 0.77$.

As a result of the *Poetic Edda* containing many loosely related texts and 4 large cliques, it has the most similar properties to the social networks of chapter 3 of all the narratives. This is somewhat surprising as this text is among the least realistic of the myths studied here. Only the disjoint degree distribution sets it apart from the social networks studied earlier.

### *5.3.3* Prose Edda

The *Prose Edda* contains 374 characters with an average degree of $\langle k \rangle = 10.45$. The beginning of the narrative deals with the creation of the world. In this, three

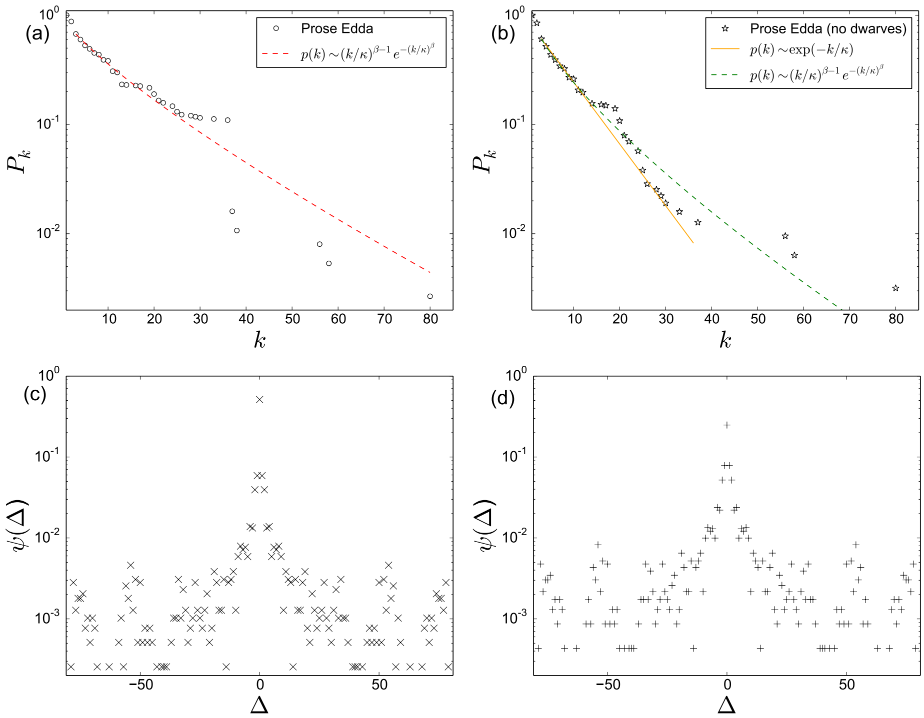



*Fig. 5.20:* Panel (a) shows the complementary cumulative degree distribution for the *Prose Edda* network. The most likely candidate model is a log normal distribution (dotted black line). There is moderate support for an exponential also (continuous yellow line). Panel (b) shows the $\psi(\Delta)$ distribution. The protagonist frequently interacts with low degree characters as seen at the ends of the plot. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the robustness of the *Prose Edda* network. It is very fragile when vertices are removed by degree or betweenness.

groups of dwarves are created and named. Here it is assumed that each dwarf knows all the other dwarves in each group. This, however, gives three large cliques of characters, none of whom appear at any point later in the text. As a result the network is studied both with and without the clique. Removing the dwarves yields 315 characters with a lower average degree of $\langle k \rangle = 7.33$. This is referred to as *Prose Edda* (no dwarves) in table 5.1 and table 5.2.

The degree distribution is shown in in fig. 5.20 (a). The most likely of the candidate models is a Weibull distribution with parameters $\beta = 0.81 \pm 0.06$, $\kappa = 10.05 \pm 0.01$. Removing the dwarves, the new degree distribution is displayed in fig. 5.20 (b). Here the parameters are $\beta = 0.74 \pm 0.05$ $\kappa = 5.7 \pm 0.02$. As shown however, without the three highest degree characters (Odin, Loki and Thor), an exponential with $\kappa = 8.54 \pm 0.03$ is also supported.

The three large cliques make the network assortative with $r_k = 0.33 \pm 0.04$, however upon removing them this drops to $r_k = 0.02 \pm 0.03$. This change can be seen in fig. 5.20 (c) and fig. 5.20 (d) where $\psi(0) = 0.53$ drops to $\psi(0) = 0.24$. The clustering assortativity also drops from $r_C = 0.52 \pm 0.03$ with the dwarves to $r_C = 0.36 \pm 0.03$ without. The Pearson similarity measure remains unchanged.

The average path length of the *Prose Edda*, $\ell = 4.36$, is longer than the average path length of the same sized random graph, $\ell_{\text{rand}} = 2.78$. Therefore the network is not small world and this is the case with or without the dwarves. The transitivity, $C_T = 0.81$, is higher than naïvely predicted, $C_{\text{n}} = 0.14$. After the removal of the clique, this drops to $C_T = 0.58$ and the naïve estimate remains unchanged.

Even without the dwarves, the *Prose Edda* still has 128 cliques, the largest of which contains 22 vertices. The community detection algorithm finds 12 communities with a modularity of $Q = 0.55$.

This network has similar properties to the social networks of chapter 3 with a high assortativity, non-power law degree distribution, high transitivity and slightly larger than small world. However the assortativity and high transitivity are the result of 59 characters introduced in three cliques at the beginning of the story. Without these, the *Prose Edda* has similar properties to other mythological networks studied here.

### *5.3.4* Völsungasaga

*Völsungasaga* is a tale spanning multiple generations of those associated with Volsung. As it takes place over such a long time period, characters cannot accumulate links throughout the entire story. Hence the maximum degree for the 103 characters is relatively low at $k_{\max} = 24$. This yields a degree distribution with a fast decaying tail as shown in fig. 5.21 (a). The most likely of the candidate models are a log-normal distribution (with parameters $\mu = 1.44 \pm 0.06$ and $\sigma = 0.8 \pm 0.03$) and a Weibull distribution (with $\beta = 0.84 \pm 0.09$ and $\kappa = 3.6 \pm 0.04$).

The network is neither assortative nor disassortative, $r_k = -0.01 \pm 0.06$ visualised in fig. 5.21 (b). This is also the case for the clustering assortativity $r_C = -0.04 \pm 0.06$. The average path length $\ell = 3.41$ is slightly longer than for a random graph $\ell_{\text{rand}} = 2.79$. The transitivity is almost 4 times larger than its naïve prediction giving it similar behaviour to the social networks in chapter 3.

The Girvan-Newman algorithm separates the friendly network into 7 communities with modularity $Q = 61$. The blue, green, red and grey nodes correspond to

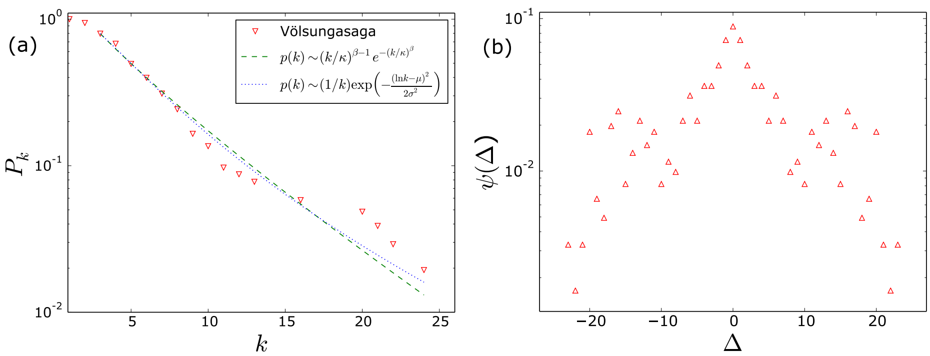



*Fig. 5.21:* Panel (a) shows the complementary cumulative degree distribution for the network of *Völsungasaga*. The most likely candidate models are a log normal distribution (dotted blue line) and a Weibull distribution (dashed green line). Panel (b) shows the $\psi(\Delta)$ distribution. The distribution highlights that there are virtually no correlations between the degree of vertices at either ends of an edge.

the four generations of characters in the story. The other coloured vertices represent different kingdoms in the text.

### 5.3.5 Nibelungenlied

The *Nibelungenlied* follows the same story as the last generation of *Völsungasaga*. Here we have only 66 vertices however, and the story goes into much more detail about the characters. This leads to a much higher average degree of $\langle k \rangle = 9.48$. The degree distribution is shown in fig. 5.23 (a). Like *Völsungasaga*, the most likely of the candidate models are a log-normal distribution, with parameters $\mu = 2.29 \pm 0.09$ and $\sigma = 0.86 \pm 0.04$, and a Weibull distribution with $\beta = 0.97 \pm 0.18$ and $\kappa = 10.79 \pm 0.03$. The $\text{AIC}_\text{c}$ and BIC weights also give mild support for an exponential distribution with $\kappa = 10.21 \pm 0.15$. The value of $\beta$ in the Weibull distribution is already very close to 1 giving further support for a pure exponential rather than a Weibull distribution.

The network is disassortative with $r_k = -0.28 \pm 0.06$. The peak of the $\psi(\Delta)$ distribution is very low indicating an absence of correlations between the degrees of interacting vertices. The network is also clustering disassortative with $r_C = -0.22 \pm 0.06$. It has a very high Pearson similarity measure of $r_P = 0.23$ highlighting that many nodes share the same neighbours.

The *Nibelungenlied* network has an average path length, $\ell = 2.14$, almost equal to its random graph, $\ell_{\text{rand}} = 2.11$. It has a high clustering coefficient,

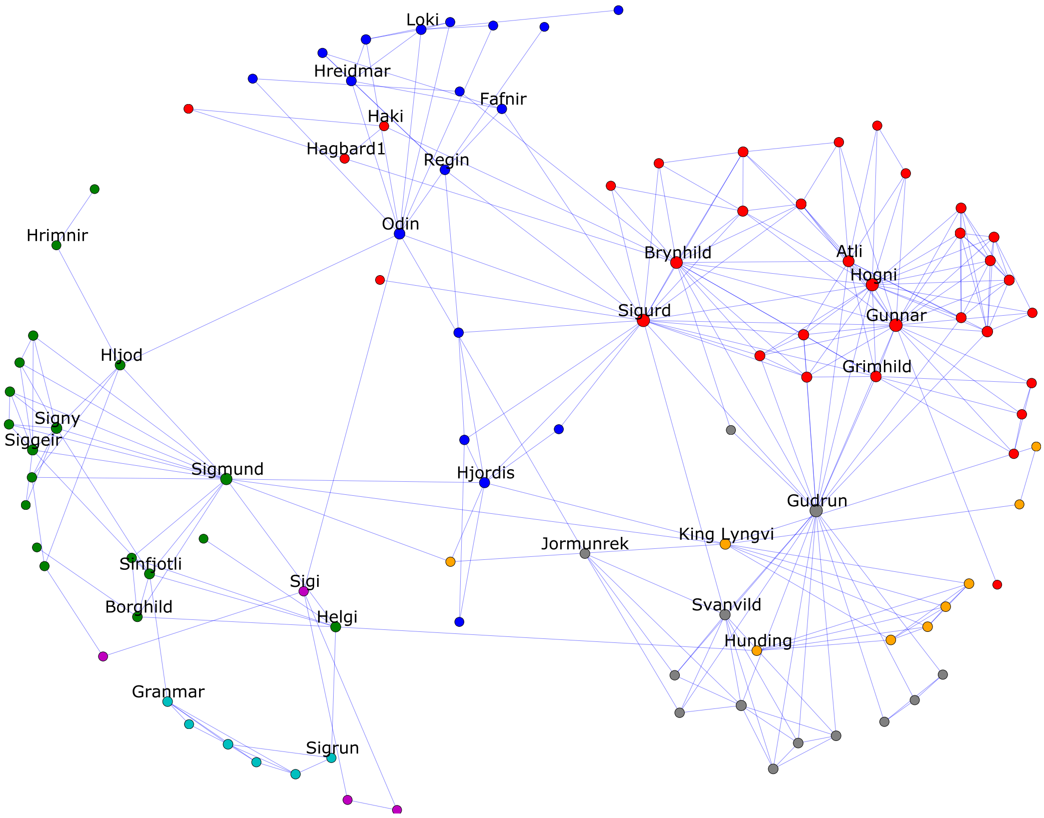



*Fig. 5.22:* The Girvan-Newman algorithm finds 7 communities in the friendly giant component of *Völsungasaga*. This separates the 4 different generations as well as some additional kingdoms in the narrative.

$C = 0.69$, however its random graph also has a relatively high clustering coefficient, $C_{\text{rand}} = 0.14$. Its transitivity, $C_T = 0.48$, is well estimated by $C_{\text{n}} = 0.51$. The network is robust to the random removal of vertices but fragile when removing nodes by degree or betweenness. The Girvan-Newman algorithm finds no community structure, the modularity never increases above $Q = 0.16$ (12 communities).

### 5.3.6 Orkneyinga saga

There are 441 unique characters in the *Orkneyinga saga*. The average degree of the network is $\langle k \rangle = 5.43$. The degree distribution is shown on a log-log scale in fig. 5.24 (a). It is best fitted by a log-normal distribution with parameters $mu = 1.31 \pm 0.03$ and $\sigma = 0.9 \pm 0.02$. However the tail of the distribution behaves a differently, when $k_{\text{min}} = 8$, a power law is the most likely model with an exponent

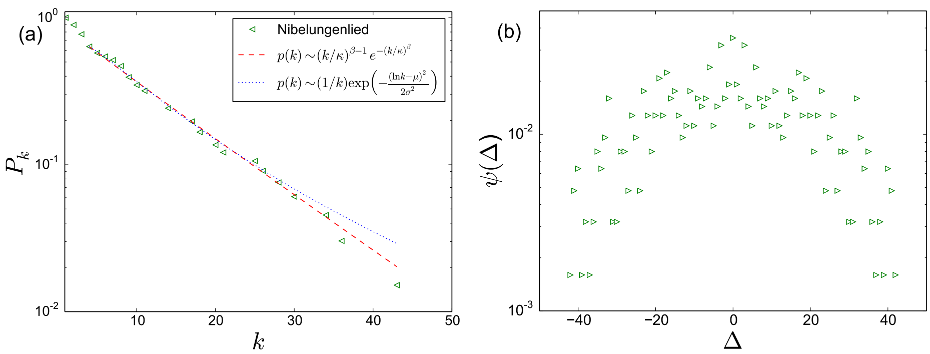



*Fig. 5.23:* Panel (a) shows the complementary cumulative degree distribution for the network of the *Nibelungenlied*. The most likely candidate models are a log normal distribution (dotted blue line) and a Weibull distribution (dashed red line). Panel (b) shows the $\psi(\Delta)$ distribution for the disassortative network.

$\gamma = 3.03 \pm 0.22$.

The network is disassortative with $r_k = -0.09 \pm 0.03$. The $\psi(\Delta)$ distribution is displayed in fig. 5.24 (b). The fluctuations for large difference in degree $\Delta$ drive the disassortativity. *Orkneyinga saga* is clustering assortative however with $r_C = 0.20 \pm 0.03$. The network has a low Pearson similarity measure, $r_P = 0.02$.

The average path length of $\ell = 5.04$ is quite high. This is in part due to the narrative containing genealogies of the Earls of Orkney which can be seen in the long diameter, $\ell_{\max} = 21$. The network also has a high clustering coefficient, $C = 0.50$, and a transitivity, $C_T = 0.27$, three times higher than predicted, $C_{\rm n} = 0.07$. The mean clustering coefficient per degree decays as a power law (with exponent $\beta = 1.04 \pm 0.08$) as shown in fig. 5.24 (c). This gives evidence of a hierarchical structure within the network.

The *Orkneyinga saga* network is fragile to the removal of vertices by degree or betweenness. It also contains 257 cliques, the largest of which comprises just 10 vertices. The average clique size is $\langle C_s \rangle = 3.96$. The clique size distribution $P_{C_s}$ is displayed in fig. 5.24 (d). It is fitted with an exponential with parameter $1.44 \pm 0.04$. The Girvan-Newman community detection algorithm finds 12 communities with modularity $Q = 0.69$.

## 5.4 *Sagas of Icelanders*

The networks of the sagas of Icelanders feature different properties depending on type of saga (i.e. family saga, outlaw saga, etc.). The sagas contain many recurring

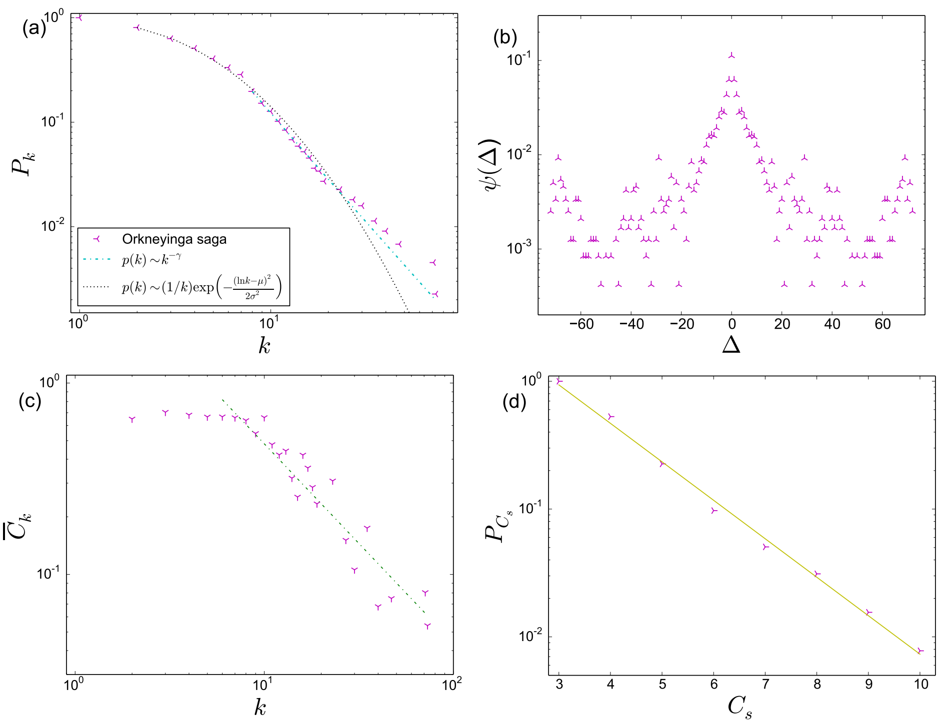



*Fig. 5.24:* Panel (a) shows the complementary cumulative degree distribution for the network of the amalgamation of *Orkneyinga saga*. The most likely candidate model is a log normal distribution (dotted black line). A power law is also shown (cyan dash-dotted line) for $k_{\text{min}} = 8$. Panel (b) shows the $\psi(\Delta)$ distribution which is disassortative. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the clique size distribution $P_{C_s}$ on a semi-log scale with a fitted exponential.

characters allowing them to be merged into a larger network. In spite of the overlap however, the sagas have differences in their network properties as outlined below and in Mac Carron & Kenna (2013a) and Mac Carron & Kenna (2013c).

### *5.4.1* Gísla saga

*Gísla saga* is the smallest of the sagas. As mentioned in section 4.3.4, this is known as an "outlaw saga" and therefore is centred on one particular character rather than a family or population. Consequently the maximum degree $k_{\text{max}} = 44$ is comparatively large next to the other sagas, for example it is just less than the most connected character in *Laxdæla saga*, which is over three times larger.

The degree distribution is shown in fig. 5.25 (a) on a log-log scale. A log-

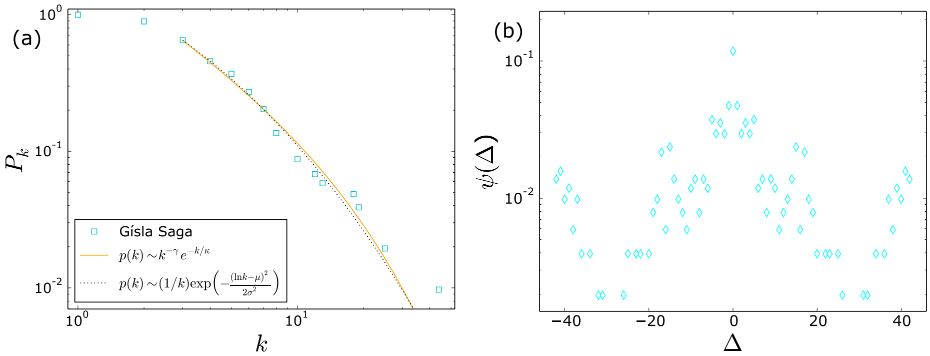



*Fig. 5.25:* Panel (a) shows the complementary cumulative degree distribution for the network of *Gísla saga*. The most likely candidate models are a log normal distribution (dotted black line) and a truncated power law (continuous orange line). Panel (b) shows the $\psi(\Delta)$ distribution.

normal distribution (with $\mu = 0.38 \pm 0.1$ and $\sigma = 1.17 \pm 0.07$) and a truncated power law (with $\gamma = 1.72 \pm 0.12$ and $\kappa = 19.08 \pm 0.02$) have the highest support. The fits do not capture the degree of the vertex of the protagonist however.

*Gísla saga* has the highest disassortativity of the sagas with $r_k = -0.15 \pm 0.04$. This is due to the high degree protagonist interacting with characters who have a large difference in degree $\Delta$ as shown in fig. 5.25 (b). The network has no clustering assortativity, $r_C = 0.01 \pm 0.07$. The Pearson similarity measure is $r_P = 0.06$ indicating a preference for characters to share the same neighbour.

The network is small world with an average path length of $\ell = 3.38$ and clustering coefficient of $C = 0.60$. Its transitivity is well predicted by its naïve estimate. The network is robust to the random removal of vertices but fragile upon their targeted removal. The community detection algorithm finds 9 communities with modularity $Q = 0.50$.

### *5.4.2* Vatnsdæla saga

The network of *Vatnsdœla saga* contains 132 vertices with an average degree of $\langle k \rangle = 4.39$. The degree distribution is shown in fig. 5.26 (a) on a log-log scale. The $\mathrm{AIC_c}$ and BIC weights give almost equal support to a truncated power law ($\gamma = 1.82 \pm 0.1$ and $\kappa = 18.88 \pm 0.02$) and a log-normal distribution ($\mu = 0.0 \pm 0.1$ and $\sigma = 1.25 \pm 0.08$).

There are no correlations between the degrees of interacting vertices, $r_k = 0.00 \pm 0.06$. This is visualised in fig. 5.26 (b) where there is no large peak about

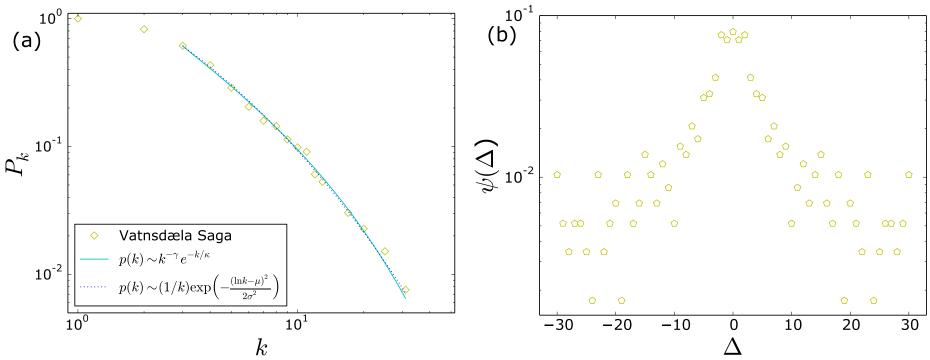



*Fig. 5.26:* Panel (a) shows the complementary cumulative degree distribution for the network of *Vatnsdæla saga*. The most likely candidate models are a truncated power law (continuous cyan line) and a log normal distribution (dotted blue line). Panel (b) shows the $\psi(\Delta)$ distribution.

the mean but also no large fluctuations at the edges. It has a low but slightly positive clustering assortativity, $r_C = 0.08 \pm 0.06$.

The network has a low average path length $\ell = 3.86$ and high clustering coefficient $C = 0.49$ making it small world. The transitivity is over three times higher than the naïve estimation. The network is fragile when removing vertices by degree or betweenness but robust to the random removal of nodes. *Vatnsdæla saga* is found to have 5 communities with modularity $Q = 0.58$. This network shares many of the features of the social networks in chapter 3.

### *5.4.3* Egils Saga

*Egils saga* is often referred to as a "poet's saga" and centres on its protagonist Egil Skallagrimsson and his travels. As mentioned in section 4.3.4, half of the saga takes place in Norway and the rest in Iceland differentiating it from the other sagas studied here. There are 292 characters in this narrative and 770 edges. Of these, 716 of the edges are friendly edges. The degree distribution is shown in fig. 5.27 (a). It is best fitted by a log-normal distribution with parameters $\mu = 1.11 \pm 0.04$ and $\sigma = 0.98 \pm 0.03$. None of the distributions fit the final few points well however.

The network is disassortative with $r_k = -0.07 \pm 0.03$. Fig. 5.27 (b) shows some large fluctuations in the difference of degrees $\Delta$ at the ends of the $\psi(\Delta)$ distribution. The friendly network however has an $r_k = -0.03 \pm 0.04$ indicating that there are no correlations between the degrees of interacting vertices. The network

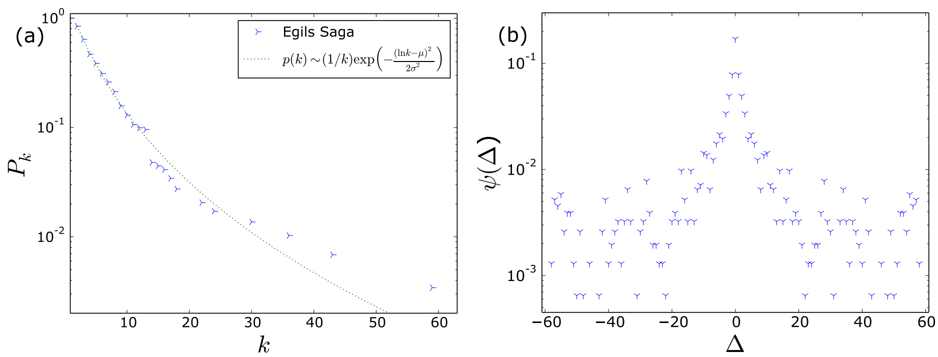



*Fig. 5.27:* Panel (a) shows the complementary cumulative degree distribution for the network of *Egils saga*. The most likely candidate model is a log normal distribution (dotted green line). Panel (b) shows the $\psi(\Delta)$ distribution.

has a high clustering assortativity with $r_C = 0.28 \pm 0.04$.

The *Egils saga* network has slightly longer average path length, $\ell = 4.19$, than the same sized random graph, $\ell_{\text{rand}} = 3.57$. Its clustering coefficient, $C = 0.56$, is much larger than the random clustering coefficient, $C_{\text{rand}} = 0.02$. The transitivity, $C_T = 0.38$, is almost 5 times higher than naïve expectation, $C_{\text{n}} = 0.08$. It is also structurally balanced with less than 4.5% of its triangles containing hostile edges.

The network is particularly fragile to the removal of vertices by betweenness. Removing 10% of the highest betweenness vertices (29 characters) leaves the giant component less than 5% connected. In the case of removal by degree it is almost 28% connected and random 92% connected. *Egils saga* contains a clique of 14 characters. Using the Girvan-Newman community detection algorithm, the modularity peaks at $Q = 0.67$ with 6 communities.

### 5.4.4 Laxdæla saga

The network of *Laxdæla saga* contains 332 vertices and an average degree of $\langle k \rangle = 5.39$. The complementary cumulative form of degree distribution is shown in fig. 5.28 (a) on a semi-log scale. It is well fitted by a log-normal distribution with parameters $\mu = 1.28 \pm 0.04$ and $\sigma = 0.91 \pm 0.02$. An exponential distribution ($\kappa = 5.63 \pm 0.02$) is also shown excluding the 4 highest degree characters.

*Laxdæla saga* has the highest assortativity of the sagas with $r_k = 0.19 \pm 0.04$. This $\psi(\Delta)$ distribution is very uniform as displayed in fig. 5.28 (b). This may be in part due to the narrative being a family saga and not centred on one particular protagonist. The network is also clustering assortative with $r_C = 0.25 \pm 0.04$.

The average path length, $\ell = 5.01$, is longer than expected compared to its random counterpart, $\ell_{\text{rand}} = 3.60$. Therefore the network is not small world, even though it has a high clustering coefficient of $C = 0.45$. The mean clustering per degree $\bar{C}_k$ is shown in fig. 5.28 (c). It is fitted with a power law with exponent $\beta = -0.99 \pm 0.14$ showing evidence of hierarchy. *Laxdœla saga*'s transitivity, $C_T = 0.41$, is much higher than naïvely expected, $C_{\text{n}} = 0.06$.

The *Laxdœla saga* network is initially robust to all forms of vertex removal as shown in fig. 5.28 (d). However upon removal of 5% of the vertices (17 characters) in order of betweenness centrality, the size of the giant component reduces to 53% of its original size. Removing the 5% of characters by degree however, the giant component is at 88% of its original size and 95% in the case of random removal. Applying the community detection algorithm, 9 communities are found with modularity $Q = 0.56$. Of these communities, 3 of them contain more than 20 vertices.

### *5.4.5* Njáls saga

*Njáls saga* is the largest of the sagas of the Icelanders with 575 characters. It contains many more hostile edges than the other sagas with 224 of the 1612 edges being purely hostile. The degree distribution is best fitted by a log-normal distribution with parameters $\mu = 0.0 \pm 0.05$ and $\sigma = 1.46 \pm 0.05$ as displayed in fig. 5.29 (a).

The overall network has almost no assortativity $r_k = 0.01 \pm 0.02$. This is visualised in fig. 5.29 (b). The friendly network however is assortative with $r_k = 0.07 \pm 0.03$. The network is also clustering assortative $r_C = 0.12 \pm 0.03$.

Like *Laxdœla saga*, the network of *Njáls saga* is not small world with its average path length $\ell = 5.14$ being longer than the random average path length, $\ell_{\text{rand}} = 3.85$. *Njáls saga* however contains an unusually long diameter of $l_{\max} = 24$ which affects the average path length. It has a high clustering coefficient, $C = 0.42$, and its transitivity is almost 3 times its naïve prediction. The network is structurally balanced with 9.7% of the triangles containing an odd number of hostile edges.

The network is most fragile when removing vertices by betweenness centrality and reasonably robust when removing nodes by degree. Despite its large size, its biggest clique contains only 11 vertices. The community detection algorithm requires the removal of 412 edges to separate into smaller components. This yields 44 communities (9 of which have more than 20 vertices) with modularity $Q =$

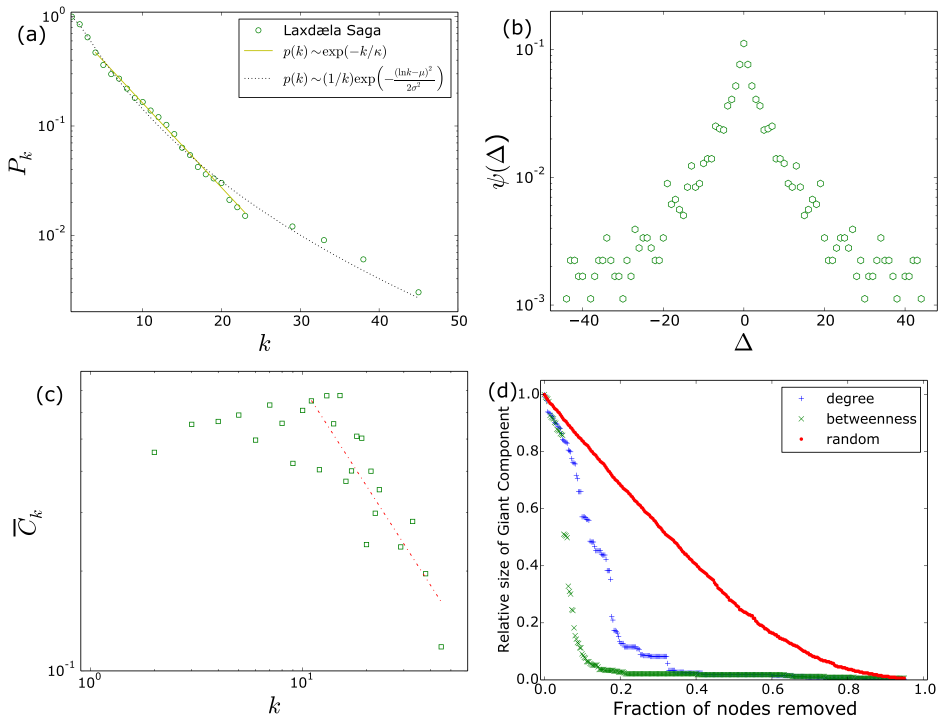



*Fig. 5.28:* Panel (a) shows the complementary cumulative degree distribution for the network of *Laxdæla saga*. The most likely candidate model is a log-normal distribution (dotted black line). There is moderate support for an exponential also (continuous yellow line). Panel (b) shows the $\psi(\Delta)$ distribution and the network is clearly assortative. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the robustness of the *Laxdæla saga* network. It is initially quite robust but becomes fragile when continually removing vertices by betweenness.

0.62.

### 5.4.6 Amalgamation of the Sagas

Initially we amalgamate the 5 larger sagas discussed up to now. The resulting network has 1282 characters and properties similar to those of Njal's saga. We are interested in breaking this network back down to see if we can separate the 5 distinct sagas.

Using the Girvan-Newman algorithm, we break the amalgamated network down until it has five large components. These have sizes 670, 230, 136, 129 and 105, which are not dissimilar to the sizes of each original network (table 5.1). Of the five emergent communities, those corresponding to *Egils saga*, *Vatnsdæla saga* and

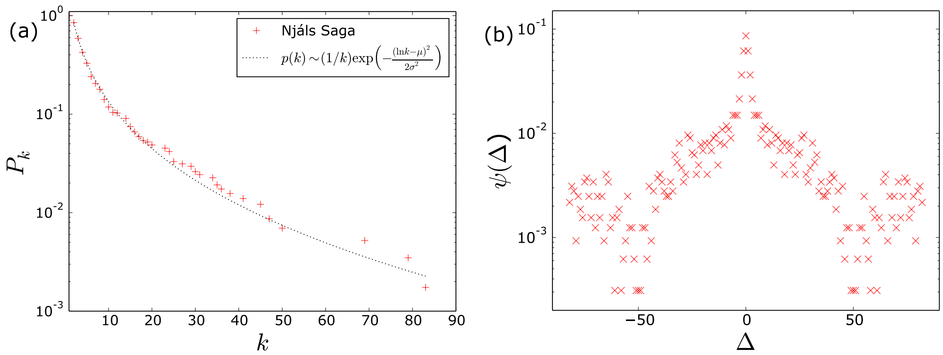



*Fig. 5.29:* Panel (a) shows the complementary cumulative degree distribution for the network of *Njáls saga*. The most likely candidate model is a log-normal distribution (dotted black line). Panel (b) shows the $\psi(\Delta)$ distribution which has an assortativity of almost 0.

*Gísla saga* emerge over 80% intact – see table 5.3. However, the breakdown to five components delivers $Q \approx 0.5$ and fails to separate the societies of *Njáls saga* and *Laxdæla saga* as, not only are there multiple characters that appear in both, but these characters often interact with different people in each narrative.

To separate *Njáls saga* and *Laxdæla saga*, one more step is required. Indeed, the modularity for the full network reaches a plateau at $n = 6$ communities with $Q \approx 0.7$. the largest component now contains 463 characters, 91% of which are from *Njáls saga*. The third largest component contains 207 characters, 80% of which are from *Laxdæla saga*. However, the latter society emerges split into two separate components.

The large overlap between *Njáls saga* and *Laxdæla saga* is visible in the network representation of Fig.5.30. In the figure, characters from each of the five

*Tab. 5.3:* Percentages of characters from different sagas which emerge when the amalgamated network is broken into 5 components. Note that the percentages can sum to more than 100 as the sagas share characters.

| Component size | Main society | Secondary society |
|---|---|---|
| 670 | 67% in *Njáls saga* | 30% in *Laxdæla Saga* |
| 230 | 85% in *Egils saga* | 15% in *Njáls Saga* |
| 136 | 82% in *Vatnsdæla saga* | 13% in *Njáls Saga* |
| 129 | 59% in *Laxdæla saga* | 51% in *Njáls Saga* |
| 105 | 85% in *Gísla saga* | 18% in *Laxdæla Saga* |

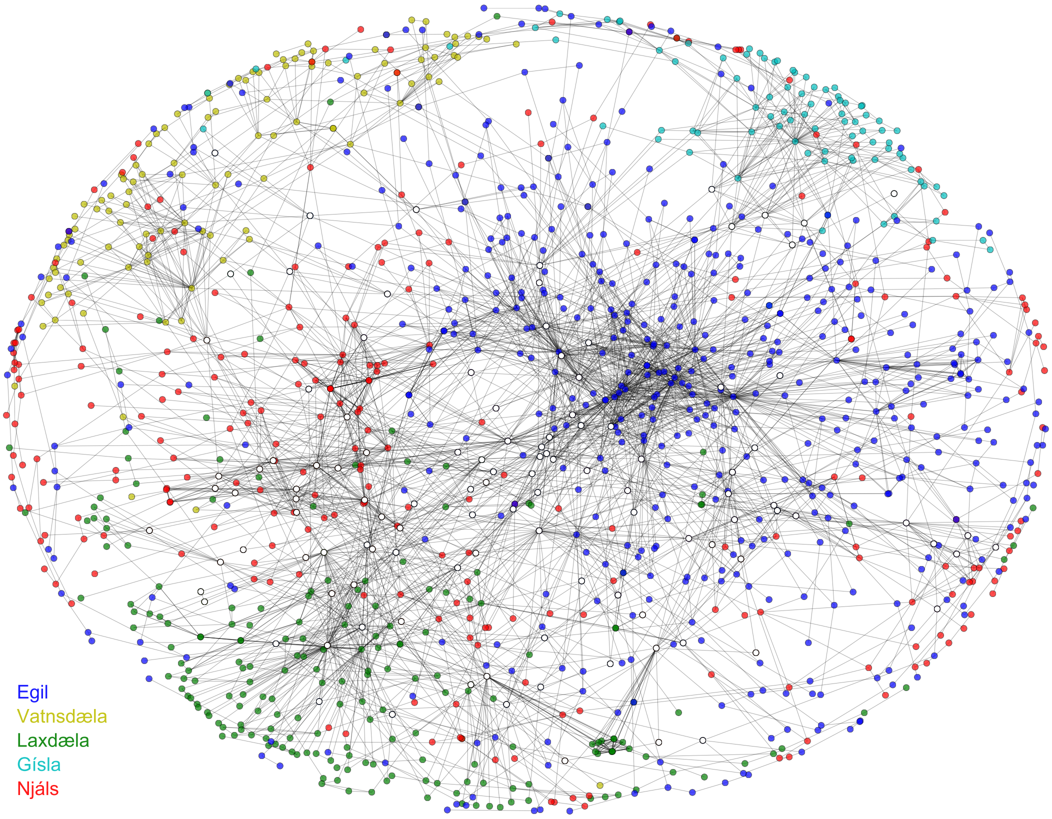



*Fig. 5.30:* Network for the amalgamation of the five major sagas. White nodes represent characters who appear in more than one saga. There is a large overlap of characters from *Laxdæla saga* (green) and *Njáls saga* (red).

major sagas are colour coded. The characters in *Laxdæla saga* appear the most scattered indicating that it is more weakly connected than some of the other sagas. It has been suggested that the author of *Njáls saga* used *Laxdæla saga* as a source (Mahnusson & Pálsson 1960; Hamer 2008), this would explain the extent of the overlap.

Next all 18 sagas are amalgamated. This network contains 1546 unique characters making it the largest network in this study. The properties of the network are most similar to that of the largest individual network, *Njáls saga*. The average degree is $\langle k \rangle = 5.52$. The degree distribution is shown in fig. 5.31 (a). It is best fitted by a log-normal distribution with $\mu = 0.70 \pm 0.02$ and $\sigma = 1.20 \pm 0.02$.

The network is assortative with $r_k = 0.06 \pm 0.02$. The $\psi(\Delta)$ distribution is displayed in fig. 5.31 (b). It is also clustering assortative $r_C = 0.16 \pm 0.02$.

The average path length $\ell = 5.58$ is longer than that of a random graph with

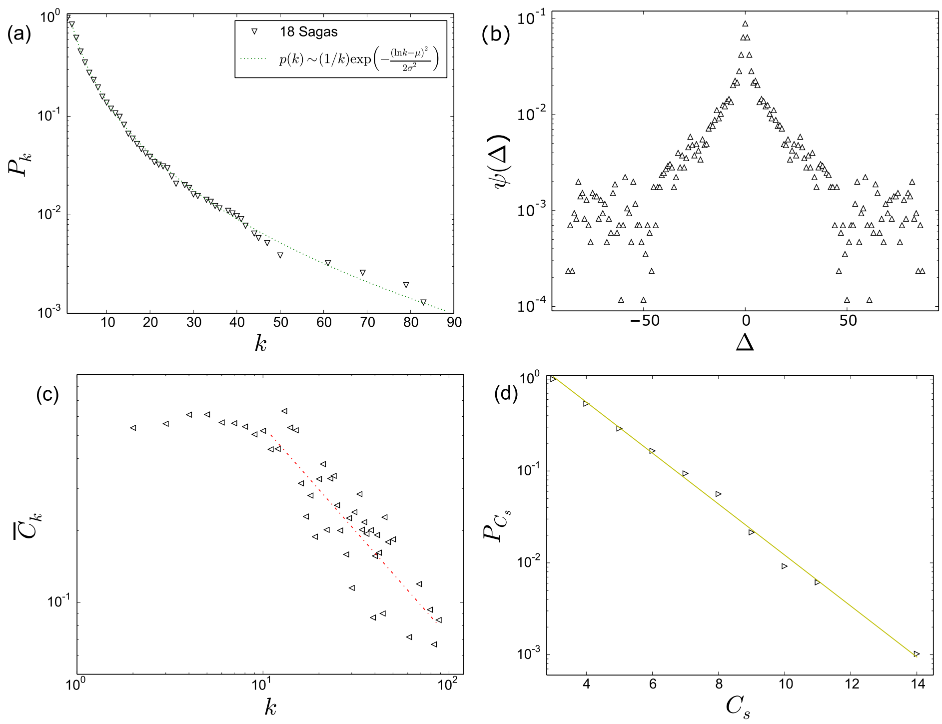



*Fig. 5.31:* Panel (a) shows the complementary cumulative degree distribution for the network of the amalgamation of 18 sagas. The most likely candidate model is a log-normal distribution (dotted green line). Panel (b) shows the $\psi(\Delta)$ distribution which is assortative. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the clique size distribution $P_{C_s}$ on a semi-log scale with a fitted exponential.

this size and average degree, $\ell_{\text{rand}} = 4.46$. The clustering coefficient, $C = 0.46$, is much larger than its random equivalent, $C_{\text{rand}} = 0.004$. The mean clustering coefficient per degree is shown in fig. 5.31 (c) with a fitted power law with exponent $\beta = 0.88 \pm 0.09$. The transitivity, $C_T = 0.28$, is much larger than the naïve estimation, $C_{\text{n}} = 0.02$. This network has the same properties as the social networks of chapter 3.

The network contains 978 cliques with an average size of $\langle C_s \rangle = 4.19$. The clique size distribution $P_{C_s}$ is displayed on a semi-log scale in fig. 5.31 (d). An exponential distribution with parameter $1.56 \pm 0.04$ is fitted to $P_{C_s}$. The modularity peaks at $Q = 0.67$ with 11 different communities. This separates the 5 larger sagas but it cannot easily break them down into their 18 components due to the high overlap between characters.

## 5.5 Welsh & Arthurian

The three Arthurian networks have very similar network properties. They are all small world, disassortative, have lower than expected transitivities and are best fitted by power or truncated power laws. The *Mabinogion*, which is a collection of shorter Welsh tales has different properties as discussed next.

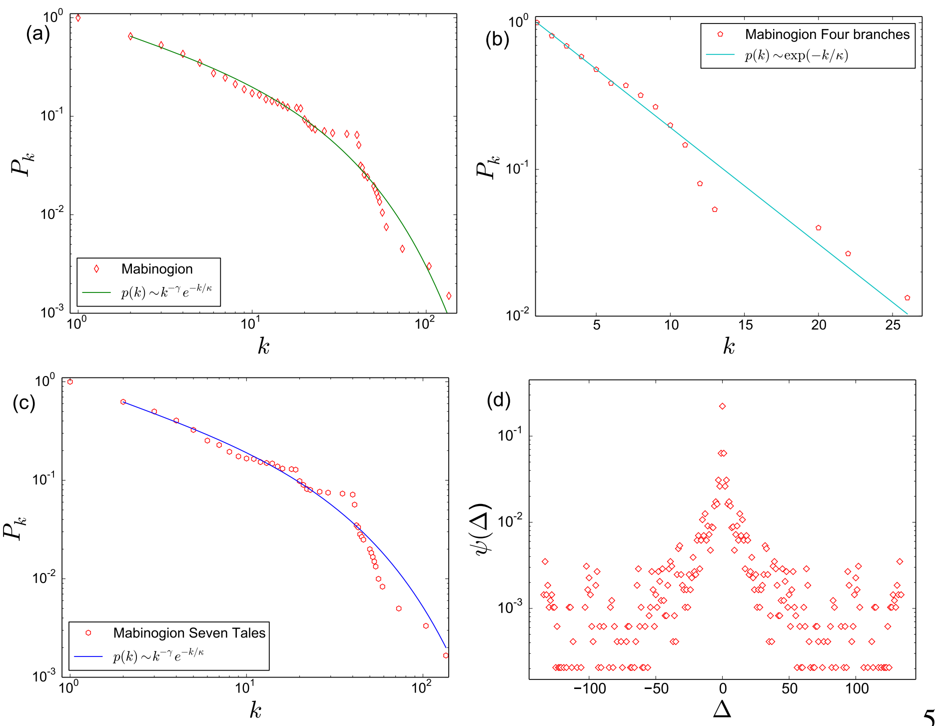
### 5.5.1 Mabinogion

The *Mabinogion* collects 11 different tales. As mentioned in section 4.3.5, there are two distinct sections however, the first four "branches" containing 75 characters and the last seven narratives with 601 characters. The properties for the overall network, the four branches and seven tales are given in table 5.1 and table 5.2. There are only 10 characters shared between the two sections giving a total of $N = 666$. The overall network of the *Mabinogion* shares similar properties to the network of the seven texts, mostly due to the larger size of the seven tales compared to the four branches. Less than 4% of the edges in the *Mabinogion* are purely hostile links.

The degree distribution for the complete network is shown in fig. 5.32 (a). It is best fitted by a truncated power law with parameters $\gamma = 1.22 \pm 0.04$ and $\kappa = 39.39 \pm 0.01$. There are two large cliques of sizes 41 and 20 that can be clearly identified. The distribution for the four branches is displayed in fig. 5.32 (b). The most likely of the candidate models is an exponential distribution with $\kappa = 5.61 \pm 0.03$. In the case of seven tales, the degree distribution is similar to that of the complete network as shown in fig. 5.32 (c). It is also best fitted by a truncated power law with $\gamma = 1.31 \pm 0.04$ and $\kappa = 55.09 \pm 0.01$.

The overall network is assortative with $r_k = 0.19 \pm 0.03$ and $r_C = 0.37 \pm 0.03$. There are two factors behind this; the first is that each narrative contains a small number of characters relative to the full network but the protagonist of one story tends not to feature in another. Therefore this does not allow a character to mass a particularly high degree compared to others. The second factor driving the assortativity is the two large cliques, the degree of each member of the larger one is higher than the average degree at the end of an edge $E[k]$. The network of the four branches however has no assortativity, $r_k = -0.03 \pm 0.06$ and $r_C = -0.00 \pm 0.07$. The assortativity of the seven tales is much the same as that of the entire network, $r_k = 0.15 \pm 0.03$ and $r_C = 0.38 \pm 0.03$.

Each *Mabinogion* network is small world with the complete network having




5

*Fig. 5.32:* Panel (a) shows the complementary cumulative degree distribution for the network of the *Mabinogion*. The most likely candidate model is a truncated power law (continuous green line). The degree distribution for the four branches is shown in panel (b). This is best fitted by an exponential distribution (continuous cyan line). Panel (c) shows the degree distribution for the seven tales which, like the complete network, is fitted with a truncated power law. Panel (d) shows the $\psi(\Delta)$ distribution. The network is assortative.

an average path length of $\ell = 3.48$ and clustering coefficient $C = 0.48$. The transitivity is higher than the naïve prediction in each case too. Each network is robust to the random removal of vertices but fragile when targeting nodes by degree or betweenness centrality.

The Girvan-Newman algorithm completely separates the two sections of the *Mabinogion* with the removal of 25 edges. This yields a modularity of $Q = 0.26$ and also splits each section in two. The modularity peaks at $Q = 0.47$ with 13 communities. This however does not split up the 11 stories as some of the Arthurian ones are highly interconnected.

The overall *Mabinogion* network has similar properties to those of the social networks in chapter 3. One distinction however, is that it has a power-law regime in its degree distribution. Like the *Poetic Edda* earlier, as this network is made up

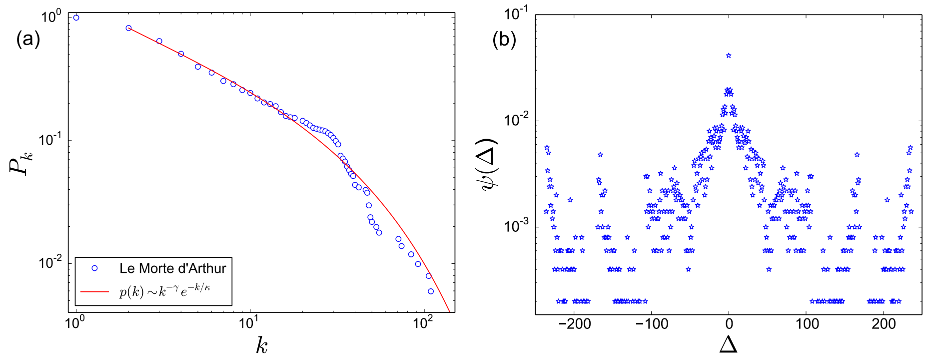



*Fig. 5.33:* Panel (a) shows the complementary cumulative degree distribution for the network of *Le Morte d'Arthur*. The most likely candidate model is a truncated power law (continuous red line). Panel (b) shows the $\psi(\Delta)$ distribution. The large fluctuations at the extremities of the figure make the network disassortative.

of many loosely connected narratives and contains some large cliques.

### *5.5.2* Le Morte d'Arthur

*Le Morte d'Arthur* is the largest and most comprehensive of the Arthurian narratives studied here. It contains 504 characters and has a relatively high average degree of $\langle k \rangle = 9.91$. The degree distribution is shown in fig. 5.33 (a). It is best fitted by a truncated power law with parameters $\gamma = 1.39 \pm 0.03$ and $\kappa = 73.89 \pm 0.01$.

In fig. 5.33 (b), we observe frequent differences of degrees with $\Delta > 200$. This the network is disassortative with $r_k = -0.230.02$. It is also clustering disassortative with $r_C = -0.04 \pm 0.02$. This network however has a relatively high Pearson similarity measure $r_P = 0.09$ indicating that vertices commonly share the same neighbours.

The network is small world, with a low average path length of $\ell = 2.76$ and a clustering coefficient of $C = 0.59$. However its transitivity is lower than expected. The network is fragile to the targeted removal of vertices but robust upon random removal.

*Le Morte d'Arthur* has 698 cliques, the largest of which contains 27 vertices. The average clique size is relatively high with $\langle C_s \rangle = 6.44$. The community detection algorithm fails to ever reach a modularity above $Q = 0.26$. This finds 28 communities however the largest of which has 329 vertices.

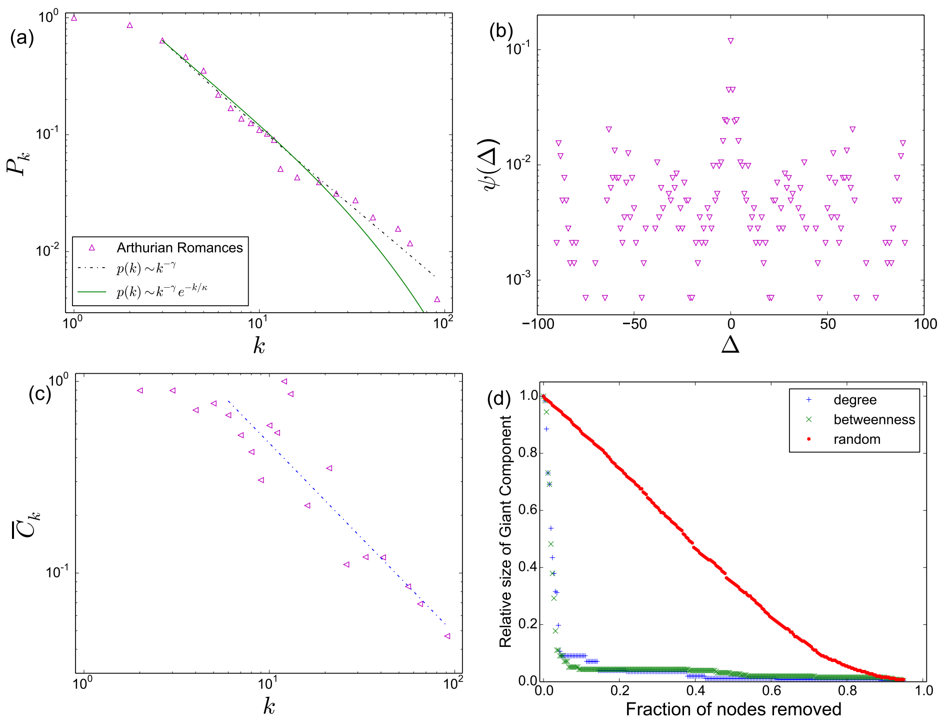



*Fig. 5.34:* Panel (a) shows the complementary cumulative degree distribution for the network of *Arthurian Romances*. The most likely candidate model is a truncated power law (continuous green line). There is moderate support for a power law also (black dot-dashed line). Panel (b) shows the $\psi(\Delta)$ distribution and the network is clearly disassortative. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the robustness of the *Arthurian Romances* network. It is unusually fragile to the targeted removal of vertices.

### *5.5.3* Arthurian Romances

The network of *Arthurian Romances* has 255 characters and an average degree of $\langle k \rangle = 5.60$. The network contains only 33 purely hostile edges. The degree distribution is shown on a log-log scale in fig. 5.34 (a). The $AIC_c$ and BIC weights give best support for a truncated power law with parameters $\gamma = 2.10 \pm 0.05$ and $\kappa = 91.59 \pm 0.01$ but also give support for a power law with exponent $\gamma = 2.30 \pm 0.10$.

As well as the degree distribution, this network has similar properties to *Le Morte d'Arthur*. It is also degree and clustering disassortative; $r_k = -0.29 \pm 0.03$ and $r_C = -0.120.04$. The $\psi(\Delta)$ distribution has many high peaks for large difference in degree $\Delta$ as shown in fig. 5.34 (b). The network also has a higher than normal Pearson similarity measure $r_P = 0.08$.

The *Arthurian Romances* network is small world with an average path length of $\ell = 2.92$ and a clustering coefficient of $C = 0.67$. The network has a lower transitivity, $C_T = 0.19$, than naïvely predicted, $C_{\mathrm{n}} = 0.30$. The mean clustering coefficient per degree $\bar{C}_k$ is displayed in fig. 5.34 (c). It is shown with a fitted power law with exponent $\beta = -0.99 \pm 0.13$.

The network is more fragile than usual. The removal of the top 5% of characters (13 vertices) by either degree or betweenness results in the giant component shrinking to less than 10% of its original size. This may be in part due to the text containing five separate narratives. To investigate this further, we use the Girvan-Newman algorithm to try to separate the different narratives. The modularity peaks at $Q = 0.56$ with 6 communities. This completely separates the first three tales from the final two, however these two contain many overlapping characters and so are not entirely separated. Instead 3 different communities are found within the amalgamation of the final two texts.

### *5.5.4* Queste del Saint Graal

*Queste del Saint Graal* is the smallest of the three Arthurian networks with 122 characters, however it has similar network properties. Of the 225 edges only 6 are purely hostile. The degree distribution is shown in fig. 5.35 (a). A truncated power law ($\gamma = 2.05 \pm 0.08$ and $\kappa = 36.59 \pm 0.02$) and a power law ($\gamma = 2.34 \pm 0.15$ both receive strong support from the $\mathrm{AIC_c}$ and BIC weights though the former is slightly favoured.

The network is disassortative $r_k = -0.23 \pm 0.05$ and fig. 5.35 (b) shows that there are many interacting vertices with large differences in degree $\Delta$. Unlike the previous two Arthurian networks however, *Queste del Saint Graal* is clustering assortative $r_C = 0.11 \pm 0.07$. The Pearson similarity $r_P = 0.08$ is relatively high again however.

The average path length, $\ell = 3.35$, is smaller than its random graph, $\ell_{\mathrm{rand}} = 3.74$, and its clustering coefficient, $C = 0.45$, is larger than $C_{\mathrm{rand}} = 0.03$. The transitivity of $C_T = 0.22$ is well predicted, though slightly smaller, than its naïve estimate of $C_{\mathrm{n}} = 0.27$. The network is robust to the random removal of vertices but fragile when nodes are removed by degree or betweenness. Applying the community detection algorithm, the network is found to have 12 communities with modularity $Q = 0.47$.

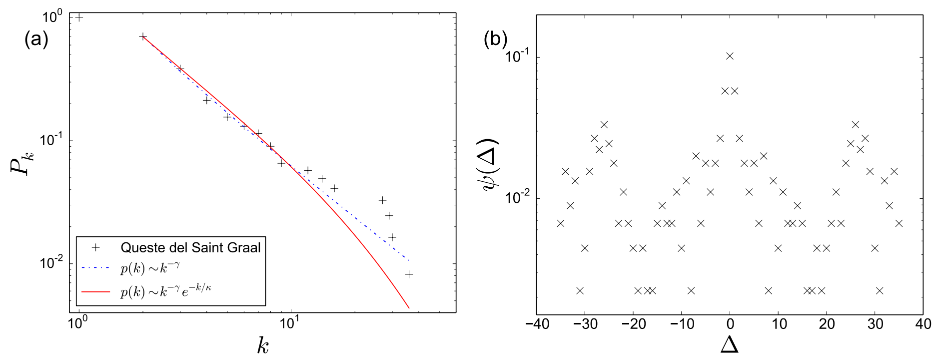



*Fig. 5.35:* Panel (a) shows the complementary cumulative degree distribution for the network of *Queste del Saint Graal*. The most likely candidate models are a truncated power law (continuous red line). Panel (b) shows the $\psi(\Delta)$ distribution. The large fluctuations at the extremities of the figure make the network disassortative.

### *5.5.5* Tristan

The *Tristan* network is one of the smaller networks analysed here. It contains only 49 characters with an average degree of $\langle k \rangle = 5.43$. Due to its small size, it is hard to determine the functional form of its degree distribution. The $\mathrm{AIC_c}$ and BIC weights offer support for each of the candidate models. A log-normal distribution with parameters $\mu = 0.49 \pm 0.15$ and $\sigma = 1.23 \pm 0.12$ is slightly favoured and is shown in fig. 5.36 (a). A power law with exponent $\gamma = 2.23 \pm 0.22$ is also shown however as this is the only fit to predict the vertex with the highest degree.

The network is disassortative $r_k = -0.37 \pm 0.05$. This is visualised in fig. 5.36 (b). *Tristan* is also clustering disassortative $r_C = -0.33 \pm 0.09$ and has a high Pearson similarity $r_P = 0.28$ due to vertices sharing the same neighbours. As with the Arthurian networks, *Tristan* is small world with an average path length of $\ell = 1.99$ and clustering coefficient of $C = 0.75$. Similarly its transitivity, $C_T = 0.27$, is lower than predicted, $C_{\mathrm{n}} = 0.61$.

This network is most fragile when vertices are removed in order of betweenness but only robust when nodes are randomly removed. Once again the small size of the network makes it difficult to find community structure. The largest clique contains 6 characters and the modularity with the Girvan-Newman algorithm never goes above $Q = 0.25$, peaking with 13 communities, only two of which contain more than 5 characters.

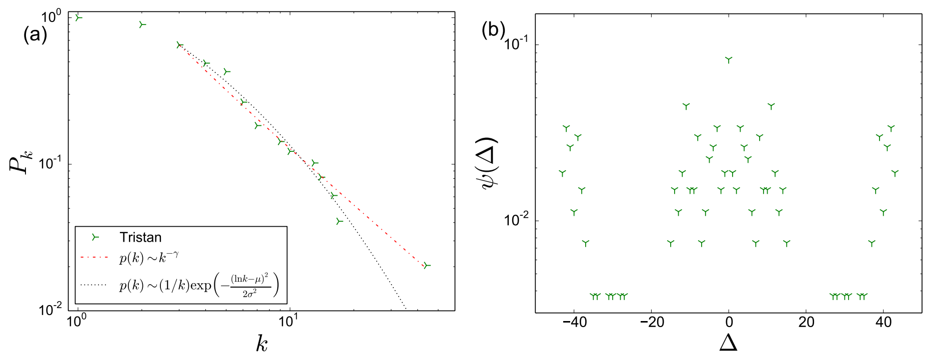



*Fig. 5.36:* Panel (a) shows the complementary cumulative degree distribution for the network of *Tristan*. All of the candidate models are offered some support but a log-normal distribution is slightly favoured (dotted black line). A power law is also shown (dash-dotted red line). Panel (b) shows the $\psi(\Delta)$ distribution. The large fluctuations at the extremities of the figure make the network disassortative.

## 5.6 World Mythology

Of the three non-European myths studied here, two of them come from the Americas. They are both small world and disassortative but their degree distributions are somewhat different as portrayed below.

### 5.6.1 Epic of Gilgamesh

The *Epic of Gilgamesh* is the oldest and smallest of all the narratives studied here. It contains 46 characters and just 81 edges (2 of which are hostile). Like *Tristan* above, its size makes it difficult to evaluate its degree distribution. Support is offered for each of the candidate models with a slight favouring towards a log-normal distribution ($\mu = 0.0 \pm 0.18$ and $\sigma = 1.41 \pm 0.18$) and a truncated power law ($\gamma = 1.57 \pm 0.16$ and $\kappa = 19.58 \pm 0.03$) as shown in fig. 5.37 (a).

The network is disassortative with $r_k = -0.34 \pm 0.10$. The $\psi(\Delta)$ distribution shown in fig. 5.37 (b) displays an absence of any structure. *Gilgamesh* has no significant clustering assortativity, $r_C = 0.10 \pm 0.11$, but has a high Pearson similarity, $r_P = 0.11$. The network is small world with an average path length of $\ell = 2.54$ and clustering coefficient of $C = 0.46$. Its transitivity, $C_T = 0.27$, is well estimated by $C_{\rm n} = 0.29$.

The *Gilgamesh* network is robust to the random removal of vertices but upon removing the two protagonists, King Gilgamesh and Enkidu, the giant component

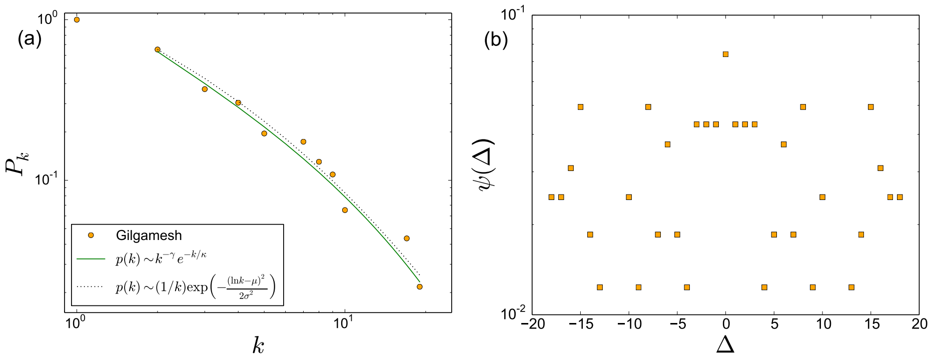



*Fig. 5.37:* Panel (a) shows the complementary cumulative degree distribution for the network of the *Epic of Gilgamesh*. All of the candidate models are offered some support but a log-normal distribution (dotted black line) and truncated power law (dash-dotted red line) are slightly favoured. Panel (b) shows the $\psi(\Delta)$ distribution. The lack of any structure signifies that the network is disassortative.

reduces to 50% of its original size. The network contains only 18 cliques. The Girvan-Newman algorithm finds 4 communities with modularity $Q = 45$. These correspond to the different communities King Gilgamesh encounters on his travels.

### *5.6.2* Popol Vuh

The *Popol Vuh* contains 98 unique vertices. Of the 409 edges, 87 of these are hostile. The degree distribution is shown in fig. 5.38 (a). The $\text{AIC}_\text{c}$ and BIC weights give the most support to a log-normal distribution with parameters $\mu = 2.56 \pm 0.05$ and $\sigma = 0.53 \pm 0.01$ with $k_{\text{min}} = 5$. An exponential distribution with $\kappa = 7.84 \pm 0.05$ is also shown, however this is not fitted to the highest degree vertex.

Fig. 5.38 (b) shows no peak about the mean of the $\psi(\Delta)$ distribution. This is due to the network's disassortativity, $r_k = -0.32 \pm 0.05$. The *Popol Vuh* network has no clustering assortativity, $r_C = -0.05 \pm 0.06$, and a high Pearson similarity, $r_P = 0.10$. Considering 21% of the edges are hostile this network can not be considered to be structurally balanced with 25% of its triangles containing an odd number of hostile edges. This is the only network we observe no structural balance.

The average path length, $\ell = 2.80$, is slightly longer than the random average path length, $\ell_{\text{rand}} = 2.39$, for a graph of this size and average degree. The clustering coefficient, $C = 0.55$, is only about six times larger than the random clustering coefficient, $C_{\text{rand}} = 0.09$. The transitivity, $C_T = 0.42$, is about twice that of the

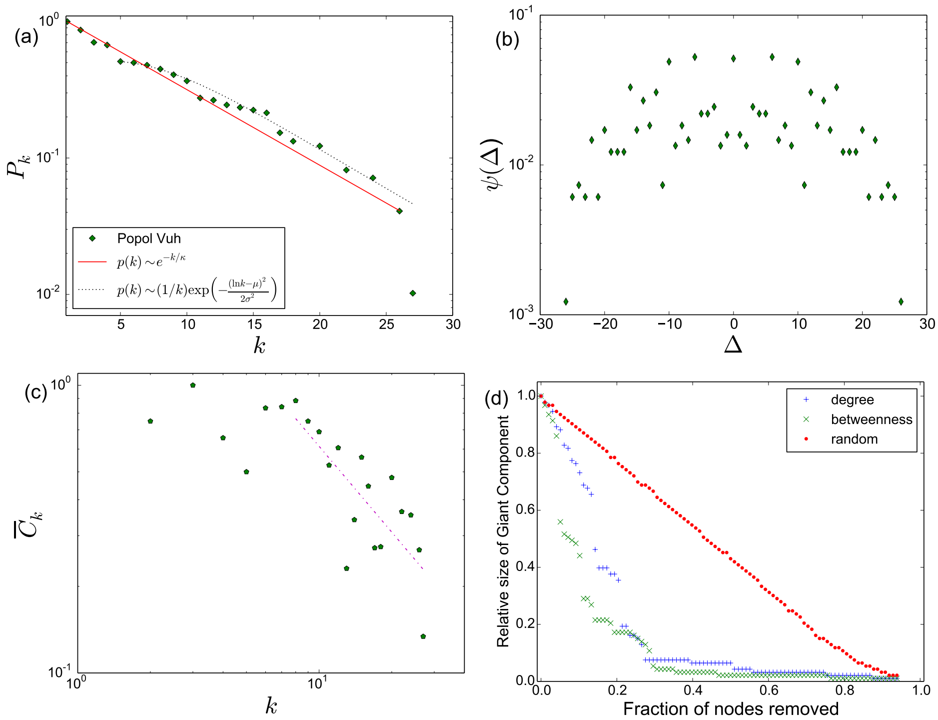



*Fig. 5.38:* Panel (a) shows the complementary cumulative degree distribution for the network of *Popol Vuh*. The most likely candidate model is a log-normal distribution (dotted black line). An exponential distribution fitted without the highest degree is also shown (continuous red line). Panel (b) shows the $\psi(\Delta)$ distribution and the network is clearly disassortative. The mean clustering per degree $\bar{C}_k$ is shown with a fitted power law in panel (c). Panel (d) shows the robustness of the *PopolVuh* network. It is more robust to the targeted removal of vertices by degree than usual.

naïve value, $C_{\mathrm{n}} = 0.11$. The network also exhibits evidence of hierarchy as shown by the decay of the mean clustering coefficient per degree in fig. 5.38 (c). A power law with exponent $\beta = 0.99 \pm 0.20$ is fitted to the tail of the distribution.

The network is robust to the removal of vertices by degree as displayed in fig. 5.38 (d). Removing the 10 highest connected vertices reduces the giant component at 77% of its original size. Randomly removing ten vertices an average of 30 times leaves the giant component at 93% of its original size. However when removing the 10 vertices with highest betweenness the network is less than 47% connected. The average clique size of the network is $\langle C_s \rangle = 5.07$ and largest clique contains 8 vertices. Using the community detection algorithm, 5 communities are found with modularity $Q = 0.53$.

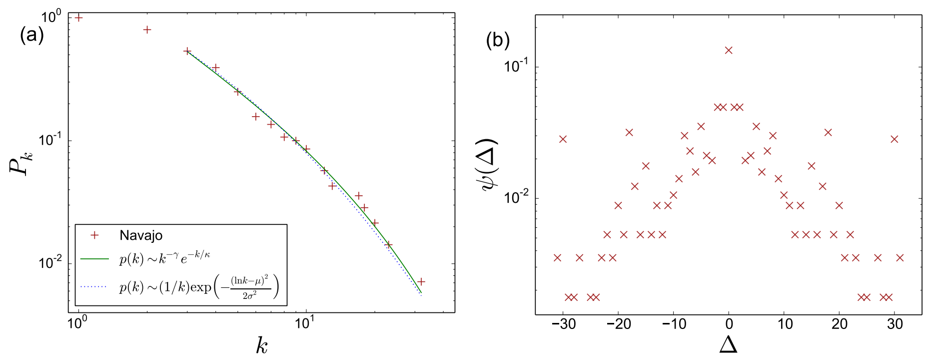



*Fig. 5.39:* Panel (a) shows the complementary cumulative degree distribution for the network of the *Navaho Indian Myths*. Both log-normal distribution (dotted black line) and truncated power law (continuous green line) are supported by the $AIC_c$ and BIC weights. Panel (b) shows the $\psi(\Delta)$ distribution. The large fluctuations at the extremities indicate that the network is disassortative.

### *5.6.3* Navaho Indian Myths

The myths of the Navaho Indians contain 140 vertices with an average degree of $\langle k \rangle = 4.04$. The network only contains 3 purely hostile edges. The degree distribution is displayed on a log-log scale in fig. 5.39 (a). The most likely candidate models are a log-normal distribution ($\mu = 0.0 \pm 0.1$ and $\sigma = 1.24 \pm 0.08$) and a truncated power law ($\gamma = 1.9 \pm 0.1$ and $\kappa = 22.36 \pm 0.02$).

The network is disassortative with $r_k = -0.18 \pm 0.05$. In fig. 5.39 (b) the $\psi(\Delta)$ distribution is shown for the difference in degree $\Delta$. The fluctuations at the extremities give rise to the disassortativity. The network is clustering assortative however with $r_C = 0.31 \pm 0.06$. It has a low Pearson similarity measure, $r_P = 0.03$.

The network is small world with an average path length of $\ell = 3.81$ and a clustering coefficient of $C = 0.44$. The transitivity, $C_T = 0.26$, is higher than the expected $C_n = 0.11$. The network is fragile when removing vertices in order of degree or betweenness but robust to random removal. Using the Girvan-Newman algorithm the network is found to have 5 communities with modularity $Q = 0.60$.

## *5.7 Fictional Networks*

Finally we turn our attention to the non-mythological networks. Like the fictional networks in chapter 3, both of these are disassortative, small world and have tran-

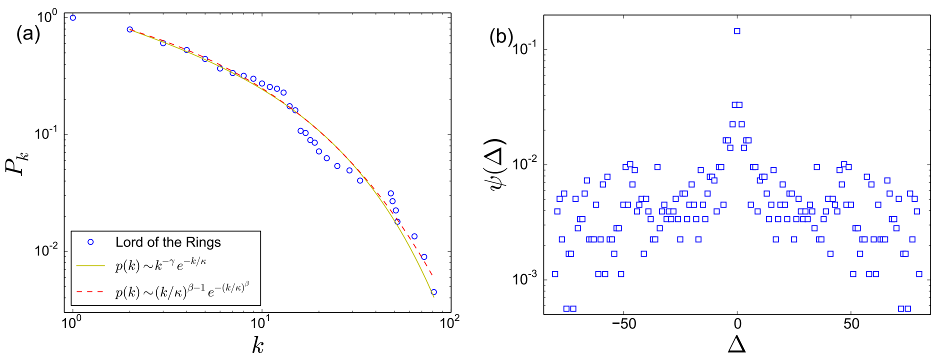



*Fig. 5.40:* Panel (a) shows the complementary cumulative degree distribution for the network of the *Lord of the Rings*. A Weibull distribution (dashed green line) and truncated power law (continuous yellow line) are equally supported by the $\mathrm{AIC_c}$ and BIC weights. Panel (b) shows the $\psi(\Delta)$ distribution for the disassortative network.

sitivity values well predicted.

### 5.7.1 Lord of the Rings

*Lord of the Rings* contains 223 characters and has an average degree of $\langle k \rangle = 7.97$. Both a truncated power law ($\gamma = 1.09 \pm 0.07$ and $\kappa = 28.45 \pm 0.01$) and a Weibull distribution ($\beta = 0.50 \pm 0.03$ and $\kappa = 2.59 \pm 0.04$) receive equal support from the $\mathrm{AIC_c}$ and BIC weights. These are shown in fig. 5.40 (a).

The network is disassortative with $r_k = -0.21 \pm 0.03$ and has no clustering assortativity, $r_C = 0.01 \pm 0.04$. The $\psi(\Delta)$ distribution is shown in fig. 5.40 (b) with long and high tails. The Pearson similarity is $r_P = 0.09$ indicating characters share neighbours.

*Tab. 5.4:* The network properties of two further fictional networks.

| | $N$ | $L$ | $\langle k \rangle$ | $k_{\max}$ | $\langle k^2 \rangle$ | $r_k$ | $r_C$ | $r_P$ |
|---|---|---|---|---|---|---|---|---|
| *Lord of the Rings* | 223 | 889 | 7.97 | 81 | 194.9 | -0.21 | 0.01* | 0.09 |
| *Dragon Tattoo* | 130 | 332 | 5.11 | 53 | 79.7 | -0.24 | 0.06* | 0.12 |

| | $\ell$ | $\ell_{\text{rand}}$ | $\ell_{\max}$ | $C$ | $C_{\text{rand}}$ | $C_T$ | $C_{\text{n}}$ | $G_C$ |
|---|---|---|---|---|---|---|---|---|
| *Lord of the Rings* | 2.90 | 2.83 | 7 | 0.63 | 0.04 | 0.37 | 0.31 | 94.6% |
| *Dragon Tattoo* | 2.88 | 3.13 | 7 | 0.57 | 0.04 | 0.28 | 0.32 | 95.4% |

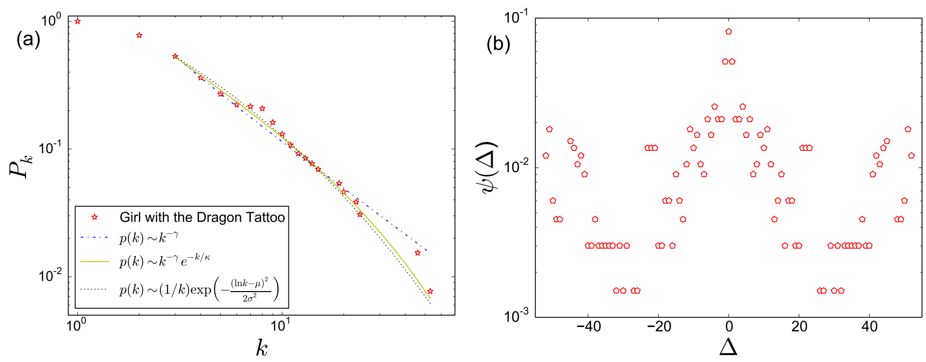



*Fig. 5.41:* Panel (a) shows the complementary cumulative degree distribution for the network of the *The Girl with the Dragon Tattoo*. A truncated power law (continuous yellow line) is the most likely of the candidate distributions. However a log-normal distribution (dotted black line) and a power law (dash-dotted blue line) are also offered support. Panel (b) shows the $\psi(\Delta)$ distribution for the disassortative network.

The network is small world and its transitivity, $C_T = 0.37$, is well estimated by $C_{\rm n} = 0.31$. The network is surprisingly robust initially to the removal of vertices both by degree and betweenness, an uncommon feature in the graphs analysed so far. Upon removal of the top 5% of characters (11 vertices) by betweenness the giant component is still at 45% of its original size. This is 63% when removing by degree and 93% when removing randomly. Though it is still susceptible to attack by betweenness centrality, most other networks are less than 20% connected upon removal of the initial 5% of characters.

### *5.7.2* The Girl with the Dragon Tattoo

The network of *The Girl with the Dragon Tattoo* (referred to as "Dragon Tattoo" in table 5.4) contains 130 vertices and 332 edges. The degree distribution is displayed in fig. 5.41 (a). A truncated power law with parameters $\gamma = 1.80 \pm 0.09$ and $\kappa = 53.59 \pm 0.01$ receives the best support of the candidate models. A log-normal distribution with parameters $\mu = 0.0 \pm 0.13$ and $\sigma = 1.49 \pm 0.12$ also has high $\text{AIC}_\text{c}$ and BIC weights and a power law with exponent $\gamma = 2.16 \pm 0.14$ receives moderate support too.

The network is disassortative with $r_k = -0.24 \pm 0.04$ as can be seen in fig. 5.41 (b). Like the network of *Lord of the Rings*, it has no clustering assortativity, $r_C = 0.06 \pm 0.06$, and a high Pearson similarity, $r_P = 0.12$. It too is small world and again its transitivity of $C_T = 0.28$ is well estimated by $C_{\mathrm{n}} = 0.32$. It is also relatively robust to the removal of vertices by betweenness centrality and degree with the giant component remaining over 40% of its original size upon removal of the top 5% of characters.

Both of these fictional networks have similar properties to the two fictional networks studied in chapter 3; namely the Marvel Universe and *Les Misérables*, as well as some of the heroic epics above.

# 6. RESULTS

In the previous chapter each mythological network was analysed and grouped according to the culture it originated from. However it is clear from table 5.1 and table 5.2 that the network properties from a given culture's myths are not always the same. Here we attempt to group them based on similar network properties in a manner akin to universality classes in statistical physics and to the Aarne-Thompson system for folktales.

## 6.1 Network Properties

The network properties discussed in chapter 2 are applied to each of the mythological networks in the previous chapter. However, when comparing the different networks to one another, we find some properties are better than others when discriminating between them. Properties such as the robustness and clustering hierarchy illuminate details about specific networks but are not as useful when comparing networks.

In chapter 3, it was found that social networks are likely to have the following properties:

- Fast decaying degree distribution with no power-law regime
- Close to small world; $\ell \gtrsim \ell_{\text{rand}}$, $C \gg C_{\text{rand}}$
- Assortativity; $r_k > 0$, $r_C > 0$
- High clustering coefficient (transitivity); $C_T > C_{\text{n}}$
- Community structure.

Most of the mythological networks contain some level of community structure. This can as be hard to quantify, however, as the Girvan-Newman algorithm will eventually find communities. There may be a few exceptions where the modularity never becomes large, but since community structure is almost ubiquitous, we do not employ it as a discriminator in the classification of the networks below. Similarly as most networks here have high clustering coefficients and are clustering assortative, we omit these properties from our scheme. We therefore restrict ourselves to the average path length $\ell$ when discussing small world and the degree assortativity $r_k$ when discussing the assortativity.

Hence, we try to characterise the networks from four properties, (i) the functional form of the degree distribution, (ii) the assortativity $r_k$ (iii) the transitivity $C_T$ and (iv) the average path length $\ell$.

(i) *Degree Distribution*
In section 2.2.2, different classifications of small world networks based on their degree distribution were outlined (as in Amaral *et al.* (2000)). These are power-law degree distributions, truncated power-law distributions, and degree distributions with fast decaying tails (such as exponentials or Gaussians). However with the small empirical datasets presented here, it can be difficult to discern a strict distinction between some distributions. As a result we choose two different classifications of degree distributions; those with support for a power-law regime (truncated or otherwise) and those without support for a power-law regime from the $\text{AIC}_\text{c}$ and BIC weights.

(ii) *Assortativity*
Here we have three regimes, disassortative $r_k < 0$, no assortativity $r_k \approx 0$ and assortative $r_k > 0$. Social networks tend to be in the third but later we pool the second and third together and discuss $r_k \geq 0$.

(iii) *Transitivity*
Newman & Park (2003) show that social networks have higher transitivity than expected. While $C_\text{n}$ from eq. (2.54) works well for random graphs, it is still just an estimate. To ensure the transitivity is sufficiently higher than expected we take the criteria $C_T > 3C_\text{n}$.

(iv) *Average Path Length*
We define two regimes; small world when $\ell \lesssim \ell_\text{rand}$ and slightly larger than small world with $\ell > \ell_\text{rand}$. The social networks in chapter 3 were mostly in the latter.

## 6.2 Classification of Mythological Networks

Table 6.1 lists each of the networks in separate categories based on the four properties above. The most immediate feature distinguishing the mythological narratives is the assortativity.

### 6.2.1 Disassortative Networks

Almost all of the disassortative networks are heroic adventure stories. The only exception here is the *Orkneyinga saga*. This, however, is also one of only two disassortative networks that has a high transitivity and is not small world. Therefore we can say that each of the mythological networks that is small world, disassortative and does not have a particularly high transitivity is a "heroic adventure" narrative.

Within these we have a few further distinctions based on the degree distributions. The *Iliad* (with all edges), *Fianaigecht* and the three Arthurian romances (*Le Morte d'Arthur*, *Arthurian Romances* and *Queste del Saint Graal*) all have support for power-law regimes in their degree distributions and no support for exponentials or faster decaying distributions. Each of these features multiple protagonists as opposed to just one.

This is not the case for the majority of the disassortative networks that have no support power-law regimes. In particular, the following five networks are all well fitted by exponential distributions when excluding the highest degree vertex (or in some cases the two highest): the *Second Battle of Mag Tuired*, the *Destruction of Da Derga's Hostel*, the *Odyssey*, the *Aeneid* and *Beowulf*. Each of these has a sole protagonist, Lugh, Conare Már, Odysseus, Aeneas and Beowulf respectively. Some have a second, but less important, high degree character central to the story. Each of these five myths follow a hero's journey and end in a battle. On the other hand, the network of the *Popol Vuh* is well fitted by an exponential (fig. 5.38 (a)), but here the exponential overshoots the highest degree vertex.

A similar observation to the above five networks, can be made in the degree distributions of *Tristan* and *Gísla saga*. Although these not fitted by exponentials, the fit for the degree distribution underestimates the character with the highest degree. A particularly high degree character has implications for the assortativity. This will raise the average degree at the end of an edge $\mathrm{E}[k]$ causing the network to more likely be disassortative. This can be visualised by the $\psi(\Delta)$ distributions which have high peaks for large difference of degrees $\Delta$.

*Tab. 6.1:* An overview of the properties of the mythological networks. The assortativity is $r_k$. Degree distributions with evidence for a power-law regime receive a checkmark under $\sim k^{-\gamma}$ whereas distributions with evidence for a faster decaying tails receive a checkmark under $\sim e^{-k/\kappa}$. The transitivity is given by $C_T$ and its naïve estimate by $C_{\mathrm{n}}$. The network is not small world if the average path length $\ell$ is greater than the random average path length $\ell_{\mathrm{rand}}$.

| | $r_k$ | $\sim k^{-\gamma}$ | $\sim e^{-k/\kappa}$ | $C_T > 3C_{\mathrm{n}}$ | $\ell > \ell_{\mathrm{rand}}$ |
|---|---|---|---|---|---|
| *Iliad* (All) | $< 0$ | ✓ | – | – | – |
| *Le Morte d'Arthur* | $< 0$ | ✓ | – | – | – |
| *Arthurian Romances* | $< 0$ | ✓ | – | – | – |
| *Queste del Saint Graal* | $< 0$ | ✓ | – | – | – |
| *Fianaigecht* | $< 0$ | ✓ | – | – | ✓ |
| *Fianaigecht* (no gene) | $< 0$ | ✓ | ✓ | – | – |
| *Epic of Gilgamesh* | $< 0$ | ✓ | ✓ | – | – |
| *Navaho Indian Myths* | $< 0$ | ✓ | ✓ | – | – |
| *Tristan* | $< 0$ | ✓ | ✓ | – | – |
| *Táin Bó Cúailnge* | $< 0$ | ✓ | ✓ | – | – |
| *Orkneyinga saga* | $< 0$ | ✓ | ✓ | ✓ | ✓ |
| *Gísla saga* | $< 0$ | ✓ | ✓ | – | ✓ |
| *Mag Tuired* | $< 0$ | – | ✓ | – | – |
| *Da Derga's Hostel* | $< 0$ | – | ✓ | – | – |
| Étaín + Óengus | $< 0$ | – | ✓ | – | – |
| *Intoxication of the Ulaid* | $< 0$ | – | ✓ | – | – |
| *Odyssey* | $< 0$ | – | ✓ | – | – |
| *Aeneid* | $< 0$ | – | ✓ | – | – |
| *Beowulf* | $< 0$ | – | ✓ | – | – |
| *Nibelungenlied* | $< 0$ | – | ✓ | – | – |
| *Popol Vuh* | $< 0$ | – | ✓ | – | ✓ |
| *Egils Saga* (All) | $< 0$ | – | ✓ | ✓ | ✓ |
| *Iliad* (Friendly) | 0 | ✓ | – | – | – |
| Ulster Cycle | 0 | ✓ | ✓ | – | ✓ |
| Mythological + Ulster | 0 | ✓ | ✓ | – | ✓ |
| *Colloquy of the Ancients* | 0 | ✓ | ✓ | – | – |
| *Vatnsdœla saga* | 0 | ✓ | ✓ | ✓ | ✓ |
| *Mabinogion* 4 Branches | 0 | – | ✓ | – | ✓ |
| *Egils Saga* (Friendly) | 0 | – | ✓ | ✓ | ✓ |
| *Njáls saga* | 0 | – | ✓ | ✓ | ✓ |
| *Völsungasaga* | 0 | – | ✓ | – | ✓ |
| *Prose Edda* (no dwarves) | 0 | – | ✓ | ✓ | ✓ |
| *Mabinogion* | $> 0$ | ✓ | – | ✓ | ✓ |
| *Mabinogion* 7 Tales | $> 0$ | ✓ | – | ✓ | – |
| *Poetic Edda* | $> 0$ | ✓ | ✓ | ✓ | ✓ |
| *Bricriu's Feast* | $> 0$ | – | ✓ | – | ✓ |
| *Prose Edda* | $> 0$ | – | ✓ | ✓ | ✓ |
| 18 Sagas | $> 0$ | – | ✓ | ✓ | ✓ |
| *Laxdæla saga* | $> 0$ | – | ✓ | ✓ | ✓ |

### 6.2.2 Non-disassortative Networks

For each of the assortative networks $r_k > 0$, there is no clear protagonist and many of the narratives are collections of shorter stories. The *Mabinogion* is the only one to receive support exclusively for a power-law regime. *Bricriu's Feast* is also notable as it has $C_T < 3C_\mathrm{n}$, however none of the other networks have a higher transitivity than this ($C_T = 0.84$). While this tale features three heroes competing to be the champion, it deals with many characters from this cycle and there is no clear protagonist. In contrast to the disassortative networks, almost all of the assortative texts have a high transitivity and are not small world.

The networks that have $r_k \approx 0$ are on the border of the above protagonist distinction. Many are collections, but some have a clear protagonist (in particular: *Egils saga* (friendly network), the *Iliad* (friendly network) and the *Colloquy of the Ancients*). These, however, are among the largest of the networks. They also tend to spread over vast geographic areas focusing on different characters in each region, unlike the disassortative networks which remain with their protagonists when they move. Most of these networks are not small world.

We now pool these two categories together. Many of the networks with an assortativity $r_k \geq 0$ are collections of stories; either separate tales we have merged (such as the Selection of Ulster Cycle Myths or the amalgamation of the 18 sagas) or short stories with some recurring characters throughout (like the *Poetic Edda* or the *Mabinogion*). Due to these containing many different segments, the heroes of one tale do not get an opportunity to continue accumulating interactions as the narrative continues. As a result the difference in the degree $\Delta$ of those they interact with are not as large, or frequent, as in the heroic adventures. Similarly these texts can be broken into modules which is the reason Newman & Park (2003) cite for social networks having a high transitivity – because they contain many communities. However in the case of collections of stories the community structure is artificial.

Therefore we shift our focus to those networks with $r_k \geq 0$ that are not collections. These are the *Iliad* (friendly), the *Colloquy of the Ancients*, *Vatnsdæla saga*, *Völsungasaga*, *Egils Saga* (friendly), *Njáls saga*, *Bricriu's Feast* and *Laxdæla saga*. The first two here are the only ones that are small world and have $C_T < 3C_\mathrm{n}$. They also both have support for power-law distributions unlike most of the others. The *Colloquy of the Ancients* takes place over a very long time period and moves all over Ireland. Effectively, this is a collection of short stories that, unlike the *Mabinogion* and the *Poetic Edda*, continually deals with the same characters. This is most likely why its network properties are different to the other collections.

The friendly network of the *Iliad* features two opposing factions and contains heroes on both sides. It takes place over a short period of time in the same location. This and *Bricriu's Feast* are the only remaining texts that are not collections and not Norse. They each contain many interactions in the society and do not feature clear protagonists, instead they are stories of a people.

As observed in section 5.1.7, *Bricriu's Feast* makes clear that many of the Ulster Cycle heroes are, in fact, associated with one another. As a result this network gives a better indication of the actual social network than the other Ulster Cycle narratives, which contain the same characters but with fewer interactions between those characters. As each network deals with the same characters, this leads one to believe that the network properties are often telling us about the type of story rather than the actual social network of the society. The tales centred on heroes tend to reveal fewer interactions between the recurring characters.

If the *Edda*s are included, then almost all the Norse networks are in the same category. Each has a high transitivity, is not small world and has an assortativity $r_k \geq 0$. It is remarkable that these have similar properties when they are different types of stories. The two *Edda*s are stories of their gods, *Vatnsdæla saga*, *Laxdæla saga* and *Völsungasaga* are tales of people or a region, and *Egils Saga* and *Njáls saga* are family sagas.

### 6.2.3 Networks of Different Cultures

The Norse are not the only culture whose networks have similar properties. Each of the three Arthurian texts falls into the same category having power-law regimes, small worldness, transitivities that are not particularly high and are disassortative. Similarly, early Irish narratives from the Mythological and Ulster Cycles (with the sole exception of *Bricriu's Feast*) contain the last three of these properties. However, they rarely have power-law degree distributions, distinguishing them from the Arthurian and *Mabinogion* networks.

When the Irish myths are amalgamated, the Ulster Cycle and Fenian Cycle also have similar properties. Both have no assortativity, are not small world and have evidence for log-normal or truncated power-law degree distributions. These both contain many short stories featuring the same cast of characters. Their network properties are also distinct from the amalgamation of the sagas, or the Welsh myths of the *Mabinogion*.

The Greek and Roman myths also all have similar properties (ignoring the distinction between the full network of the *Iliad* and its friendly network). Each is disassortative, is small world and does not have a particularly high transitivity. It seems that with the exception of the Norse myths, individual myths tends to be heroic adventures.

Outside of European myths, the *Popol Vuh* is the only network that is not structurally balanced. Otherwise this network, surprisingly, has similarities to the *Gilgamesh* network. Almost all $\psi(\Delta)$ distributions contain a peak at $\Delta = 0$, and if they are disassortative further peaks for larger difference in degree $\Delta$. However, both the *Popol Vuh* and *Gilgamesh* have almost no distinct peaks and both are disassortative. This seems to distinguish them from the European heroic myths.

As the network properties allows us to discern between types of myths, this may also tell us something about the type of myths certain cultures tend towards. The Irish, Anglo-Saxon and Greek and Roman myths are mostly heroic adventure stories, an individual myth often dealing with just a single protagonist. The Welsh and American myths however are also adventure stories but with multiple protagonists. The Nordic myths are more regional stories of a people.

### 6.2.4 Comparisons to Real and Fictional Networks

If we were to make a strict comparison to the social networks of chapter 3, we would observe that only the networks in the lower section of table 6.1 have the same properties as real social networks. If we were to be a bit more lenient and allow $r_k \approx 0$, then we would find that the Nordic networks are "realistic" and that most of the disassortative networks, having the same properties as the fictional networks, are not-realistic.

Similarly it would tell us that collections of myths such as the *Mabinogion* or tales of the Irish Fianna, are more realistic. This, of course, does not imply that the society presented is real. Instead, it indicates that because there are few connections going from one tale to the next so it gives the network a community structure which is commonly observed in real social networks.

We can make adjustments to other networks, such as looking at only the friendly interactions of the *Iliad*, or reducing the degree of the most connected characters of the *Táin* (Mac Carron & Kenna 2012). If a tale such as the latter only shows a spotlight on a few characters, many interactions are lost as made clear in *Bricriu's Feast* which has many of the same characters but entirely different network properties. One must take care when making adjustments however. Reducing the degree of a highly connected character will always increase the assortativity. Caution must be taken before claims are made about the results when artificially increasing the assortativity (c.f. Miranda *et al.* (2013)). Also, a distinction must be made between 'real' and 'realistic'.

### 6.2.5 Further Considerations

An important property of many of the social networks presented here are the cliques. Large cliques (e.g. in *Bricriu's Feast*, the *Prose Edda*) have significant impact on the global properties of the network. The interactions of every pair of character within a clique increases the assortativity and transitivity, and creates a large jump in the degree distribution data. In networks such as the *Poetic Edda* and the *Iliad*, the majority of characters in the largest clique only appear once. However in *Bricriu's Feast* they are central to the entire mythology. A clique, however, is a particularly tight community, and communities are one of the key features of social networks.

Further properties of these networks that could be analysed are the role of genders in myths. Most of these myths are male-dominated but occasionally, such as in *Laxdæla saga*, the protagonist is female. This increases the proportion of male-female interactions and is higher than the other sagas (Mac Carron & Kenna 2013c). Another approach could be to analyse the evolution of the network properties as the networks grows. This could offer insights into the structures of different narratives.

It is also important to note that this selection of myths is just a sample. With the addition of further myths, particularly more non-European ones, more insights into the differences between the networks of distinct cultures could be made. An analysis of more fictional narratives from different genres could also provide further understanding of the network structure for distinct types of narratives.

## 6.3 Summary

The overview of the network properties of these myths provide us with some key insights. Network analysis distinguishes between collections of myths of a culture from a large epic. Within this we find that the myths of the Norse gods are similar to the Welsh *Mabinogion*. However they are distinguishable from their degree distributions. The collections of Irish myths from two periods have similar properties to each other and slightly different to the Norse and Welsh ones in that they have no assortativity. A collection of Navaho and South American myths also have different properties as they are disassortative.

Of the myths that are not collections, it is almost exclusively Nordic sagas that are not disassortative. The only exception is an Irish myth that has different properties to all the other Irish narratives. The majority of myths are disassortative, however within these are still distinctions. Seven of these have a significant protagonist identified by their degree distribution; namely *Destruction of Da Derga's Hostel*, the *Odyssey*, the *Aeneid*, *Beowulf*, *Tristan* and Gísla saga. These are all tales concerning a the adventure of a hero. In essence these could be seen as a universality class and similar to the Campbell (1949) notion of universality of the hero's journey. The remaining disassortative myths tend to deal with more than one protagonist. These can be distinguished from the previous myths by their degree distribution.

# 7. CONCLUSION

Many methods of comparative mythology exist, but up to now, none of them have been purely quantitative. The approach presented here not only allows us to compare the myths of different cultures to one another, but also allows us to compare different myths from the same culture to each other. This allows for unique insights to be made into the world of myth.

In total, 33 mythological social networks were analysed here. Two different types of networks are identified; the networks of individual epics and collections of myths. Within these we find further distinctions. Individual myths tend to have two categories; these are heroic adventure type tales and the stories of a people. With the exception of the Nordic narratives, most myths studied here fall into the former category. However within the heroic adventure type tales we find a further distinction between narratives that are centred on a single protagonist versus a group.

Social networks have different properties depending on how the edges are formed. We are most interested in networks in which the edges are mutual and vertices are relatively freely allowed to choose who they interact with. When gathering data for the mythological networks, edges were created on the basis that they were mutual, however, characters in these stories do not have the same freedom to choose edges as the networks in chapter 3. There is a constraint based on the focus of the story, in essence the mythological networks only ever show a spotlight of a social network centred on who the narrator is focused on at a given time. There is no information on the characters that are "offscreen" at any time and thus we encounter many very low degree vertices.

Of the mythological networks, very few have the same properties as these real social networks. If we were to compare, the ones that have these properties tend to be collections of short narratives or large regional tales of a people. However, what this comparison to real and imaginary might more likely tell us, is the nature of the type of story a culture prefers. If this is the case, it seems that the Irish, English and Americans favour the story of the super-hero "against the world"; whereas the

Norse rather the more down-to-earth tales of a people and their more large-scale conflict.

However, to simplify to the realistic versus imaginary level may take away from the subtleties and allure of the mythological narratives. After all, this is just a selection of myths from a vast global sample size. This mere drop in the ocean of myth does, however, provide us with some unique insights. It reveals that collections of myths and tales have different properties to heroic epics but similar properties to regional based narratives. The network properties allow us to make distinctions between the type of myth but more remarkably, in some cases, they allow distinguish the myths of one culture from another. But most importantly, this method provides an entire new branch in the field of comparative mythology.

# APPENDIX

# A. MAXIMUM LIKELIHOOD ESTIMATORS

As discussed in section 2.3, it has been suggested in recent years that the methods of Maximum Likelihoods yield better estimates when fitting. In the next section, this is tested for the power-law distribution. The following section gives the likelihoods to be maximised to obtain estimates for the parameters of different distributions. The final section discusses the methods of choosing the most likely of the candidate models to fit the data.

## *A.1 Power-Law Distributions*

A distribution commonly observed in complex networks is the power law (eg. Strogatz (2001); Albert & Barabási (2002)). This has the form

$$p_k \sim k^{-\gamma}, \tag{A.1}$$

where $\gamma > 1$ and $k \neq 0$. Usually the power-law behaviour is found only in the tail of the distribution beginning at some $k_{\min}$.

The normalisation of eq. (A.1) is evaluated by

$$\sum_{k=0}^{k_{\min}} \rho_k + \sum_{k=k_{\min}}^{\infty} p_k = 1, \tag{A.2}$$

where $\rho_k$ is the distribution below $k_{\min}$. Thus eq. (A.1) becomes

$$p_k = \frac{(1-\Delta)}{\zeta(\gamma, k_{\min})} k^{-\gamma} \tag{A.3}$$

where $\Delta \equiv \sum_{k=0}^{k_{\min}} \rho_k$ and $\zeta(\gamma, k_{\min}) \equiv \sum_{n=0}^{\infty} (k_{\min} + n)^{-\gamma}$ is the Hurwitz zeta function. In the case where $k_{\min} = 1$, the denominator is then just the Riemann zeta function $\zeta(\gamma)$. The cumulative distribution function is given by

$$P_k = \sum_{q=k}^{\infty} p_q = (1-\Delta)\frac{\zeta(\gamma,k)}{\zeta(\gamma,k_{\min})}. \quad \text{(A.4)}$$

This is only valid for $k \geq k_{\min}$.

In order to fit to the data, it is easier to treat the degree distribution as if it is continuous. Eq. (A.1) can then be normalised by

$$\int_0^{k_{\min}} \rho(k)dk + \int_{k_{\min}}^{\infty} p(k)dk = 1, \quad \text{(A.5)}$$

where $\rho(k)$ is the behaviour of the unknown distribution below $k_{\min}$. The normalised distribution is then

$$p(k) = (1-\delta)\frac{\gamma-1}{k_{\min}}\left(\frac{k}{k_{\min}}\right)^{-\gamma}, \quad \text{(A.6)}$$

where $\delta \equiv \int_0^{k_{\min}} \rho(k)dk$.

If this is treated for the entire data (i.e. $k_{\min} = 1$) and in the limit $\delta \ll 1$, then one observes $p(1) \approx \gamma - 1$. As the probability to have degree $k = 1$ is bound by unity, then one concludes $\gamma < 2$. Taking the first moment of eq. (A.6) yields

$$\langle k \rangle = \int_1^{\infty} kp(k) \sim \frac{1}{2-\gamma}\left[k^{2-\gamma}\right]_1^{\infty}. \quad \text{(A.7)}$$

If $\gamma \leq 2$, then we observe that $\langle k \rangle$ diverges. For the discrete case, this is just a large finite value. Empirically in these networks we observe that the mean degree is never large, usually $\langle k \rangle < 10$. As a result, we should use $k_{\min} > 1$. In general a value of $k_{\min} = 2$ is chosen here as we also wish to fit to as much data as possible. However, it is also worth considering using the modal degree as the value for $k_{\min}$ as the distribution decays beyond this point. Clauset *et al.* (2009) provide various methods for estimating $k_{\min}$, however, that study assumes large values of $N$.

We next determine the cumulative distribution function by integrating eq. (A.6),

$$P(k) = \int_k^{\infty} p(q)dq = (1-\delta)\left(\frac{k}{k_{\min}}\right)^{-\gamma+1}. \quad \text{(A.8)}$$

Thus the cumulative distribution function also follows a power law but with an exponent differing by 1 to the original. A fit to this can be used to determine an estimate for the exponent and its error.

To evaluate the exponent using MLE we take the likelihood $\mathcal{L}$ of eq. (A.6)

$$\mathcal{L}(\gamma|k) = \prod_{i=1}^{N}(1-\delta)\frac{\gamma-1}{k_{\min}}\left(\frac{k_i}{k_{\min}}\right)^{-\gamma}. \tag{A.9}$$

Taking the logarithm

$$\ln\mathcal{L} = N\ln(1-\delta) + N\ln(\gamma-1) - N\ln k_{\min} - \gamma\sum_{i=1}^{N}\ln\frac{k_i}{k_{\min}}. \tag{A.10}$$

An estimate for $\delta$ can be obtained from the data. However as $k_{\min}$ is fixed, the first term does not affect the maximisation. Eq. (A.10) can either be numerically maximised to obtain an estimate for $\gamma$ or alternatively, setting $\partial(\ln\mathcal{L})/\partial\gamma = 0$ and solving for $\gamma$ we obtain

$$\hat{\gamma} = 1 + N\left[\sum_{i=1}^{N}\ln\frac{k_i}{k_{\min}}\right]^{-1}, \tag{A.11}$$

where $\hat{\gamma}$ denotes an estimate from the data rather than the true value. If a power law is a good model for the data, then $\hat{\gamma} \approx \gamma$.

The error is calculated using eq. (2.28) to get

$$\sigma_{\hat{\gamma}} = \frac{\hat{\gamma}-1}{\sqrt{N}}. \tag{A.12}$$

However depending on how $k_{\min}$ is estimated, there can be $\mathcal{O}(1/N)$ corrections (Clauset *et al.* 2009).

As most of our datasets are quite small, however, it may not be reasonable to use these continuous approximations. The data is also rarely a complete set of discrete integers, i.e. there is often significant gaps between two consecutive degrees.

It is more accurate to look at the discrete case from eq. (A.3). Taking the $\log$ of the likelihood of this we get

$$\ln\mathcal{L} = N\ln(1-\Delta) - N\ln\zeta(\gamma, k_{\min}) - \gamma\sum_{i=1}^{N}\ln k_i. \tag{A.13}$$

This can be numerically maximised to obtain an estimate for $\gamma$. The error is calculated again by eq. (2.28) to give

$$\sigma_\gamma = \left[ N \left( \frac{\zeta''(\gamma, k_{\min})}{\zeta(\gamma, k_{\min})} - \left( \frac{\zeta'(\gamma, k_{\min})}{\zeta(\gamma, k_{\min})} \right)^2 \right) \right]^{-1/2}, \tag{A.14}$$

where $\zeta^{(m)}(\gamma, k_{\min}) = (-1)^m \sum_{n=0}^{\infty} (\ln(k_{\min} + n))^m (k_{\min} + n)^{-\gamma}$. For large $N$ eq. (A.12) gives comparable results.

Alternatively Clauset *et al.* (2009) find an approximation for the log-likelihood in eq. (A.13) as

$$\hat{\gamma} \simeq 1 - N \left[ \sum_{i=1}^{N} \ln \frac{k_i}{k_{\min} - \frac{1}{2}} \right]^{-1}. \tag{A.15}$$

### A.1.1 Comparisons of the Exponent Estimates

Denoting $\gamma_c$ as the continuous estimator for the exponent from eq. (A.11), $\gamma_d$ as the maximisation in the discrete case in eq. (A.13), $\tilde{\gamma}_d$ as the approximation for the exponent in the discrete case in eq. (A.15) and $\gamma_{ls}$ as the estimate of the exponent using a least squares fit to eq. (A.8), the different estimates for the exponent are now compared. 1,000 synthetic power laws with an exponent of $\gamma = 2.5$ were generated each containing 1,000 observations.

The synthetic power law is generated by the method in Clauset *et al.* (2009) using code provided by the author (Clauset 2007). The average exponents and their standard deviation for the 1,000 samples are shown in Table A.1. For low $k_{\min}$, the least squares fit to the cumulative distribution of eq. (A.8) gives the most accurate results $k_{\min} = 1, \gamma_{ls} = 2.47 \pm 0.19$. However as $k_{\min}$ increases, the discrete estimate from eq. (A.13) becomes the closest to the true value of 2.5.

If, instead of using the Hurwitz zeta function in eq. (A.3), we take the upper

| $k_{\min}$ | $\gamma_c$ | $\gamma_d$ | $\hat{\gamma}_d$ | $\gamma_{ls}$ |
|---|---|---|---|---|
| 1 | 1.65 ± 0.07 | 1.96 ± 0.03 | 1.77 ± 0.02 | 2.47 ± 0.19 |
| 2 | 3.07 ± 0.26 | 2.55 ± 0.08 | 2.41 ± 0.06 | 2.47 ± 0.23 |
| 3 | 2.77 ± 0.32 | 2.52 ± 0.10 | 2.47 ± 0.09 | 2.46 ± 0.23 |
| 4 | 2.66 ± 0.42 | 2.51 ± 0.14 | 2.49 ± 0.13 | 2.42 ± 0.27 |

*Tab. A.1:* Comparison of estimates for the scaling exponent of a power law distribution with an exponent of 2.5. The error is the standard deviation from the difference in estimates for 1,000 datasets. For low $k_{\min}$, $\gamma_{ls}$ is the closest and as $k_{\min}$ increases $\gamma_d$ becomes the most accurate.

limit of the power law as $k_{\max}$ as in the maximum value as suggested by Edwards *et al.* (2007), then estimates for $\gamma_d$ become some slightly less accurate in the third decimal place. Therefore we sum to some arbitrarily large value instead of $k_{\max}$.

From this we observe that the least-squares approach yields adequate results. However its errors are hard to estimate (Clauset *et al.* 2009). Therefore eq. (A.3) is be used for the MLE estimates when dealing with power laws.

## A.2 Estimating the Parameters with MLEs

In order to obtain estimates for the parameters of the model distributions in section 2.2, the log-likelihood must be numerically maximised. For a power law this is given by eq. (A.13). When depicting histograms for degree distributions, the complementary cumulative distribution function is used. For the power law, this is given by eq. (A.4). Note that this is only valid for $k > k_{\min}$. From the data we can use $P(k_{\min}) = 1 - \Delta$ to get an estimate for $\Delta$.

The following sections give the log-likelihood that is to be maximised for each model distributions in section 2.2. These are used in the code developed for fitting the degree distributions.

### A.2.1 Exponential Distributions

The normalised exponential distribution, eq. (2.22) is given by

$$p_k = (1 - \Delta) \left( \frac{1 - e^{-1/\kappa}}{e^{-k_{\text{min}}/\kappa}} \right) e^{-k/\kappa}, \tag{A.16}$$

where again $\Delta \equiv \sum_{k=0}^{k_{\text{min}}} \rho_k$ and $\rho_k$ is the distribution below $k_{\text{min}}$. The log-likelihood corresponding to this is then

$$\ln \mathcal{L} = N \ln(1 - \Delta) + N \ln \left(1 - e^{-1/\kappa}\right) - \frac{1}{\kappa} \sum_{i=1}^{N} (k_i - k_{\text{min}}). \tag{A.17}$$

Maximising this, one obtains the MLE

$$\hat{\kappa} = \left[ \ln \left( 1 + \frac{N}{\sum_{i=1}^{N} k_i - N k_{\text{min}}} \right) \right]^{-1}. \tag{A.18}$$

In the case where the exponential is fitted to the entire distribution and not just the tail, then $k_{\min} = 0$, eq. (A.18) simplifies to

$$\hat{\kappa} = \left[ \ln \left( 1 + \frac{1}{\langle k \rangle} \right) \right]^{-1}. \tag{A.19}$$

Using the leading order term from the Taylor expansion $\ln(1 + 1/x) \approx 1/x$, then for large $\langle k \rangle$

$$\kappa \approx \langle k \rangle. \tag{A.20}$$

The cumulative distribution function

$$P(k) = (1 - \Delta) e^{-(k - k_{\min})/\kappa}. \tag{A.21}$$

### A.2.2 Truncated Power Laws

Normalising the truncated power law from eq. (2.21) yields

$$p(k) = (1 - \Delta) \frac{e^{k_{\min}/\kappa}}{Z(k_{\min})} k^{-\gamma} e^{-k/\kappa}, \tag{A.22}$$

where $Z(x) \equiv \sum_{m=0}^{\infty} (x + m)^{-\gamma} e^{-m/\kappa}$. The log-likelihood is then given by

$$\ln \mathcal{L} = N \ln(1 - \Delta) + \frac{N k_{\min}}{\kappa} - N \ln Z(k_{\min}) - \sum_{i=1}^{N} \left( \gamma \ln k_i + \frac{k_i}{\kappa} \right), \tag{A.23}$$

which must be maximised numerically. The cumulative distribution is then given by

$$P(k) = (1 - \Delta) \frac{Z(k)}{Z(k_{\min})} e^{-(k - k_{\min})/\kappa}. \tag{A.24}$$

### A.2.3 Stretched Exponential and Weibull Distributions

The stretched exponential is given by

$$p(k) \sim e^{-(k/\kappa)^{\beta}}, \tag{A.25}$$

The stretched exponential is the complementary cumulative function of the Weibull distribution

$$p(k) \sim \left(\frac{k}{\kappa}\right)^{\beta-1} e^{-(k/\kappa)^{\beta}}. \tag{A.26}$$

The log-likelihood for the stretched exponential is

$$\ln \mathcal{L} = N \ln(1-\Delta) - N \ln \left( \sum_{m=k_{\min}}^{\infty} e^{-(m/\kappa)^{\beta}} \right) - \sum_{i=1}^{N} \left(\frac{k_i}{\kappa}\right)^{\beta}, \tag{A.27}$$

and for the Weibull distribution,

$$\begin{aligned} \ln \mathcal{L} = N \ln(1-\Delta) - N \ln & \left( \sum_{m=k_{\min}}^{\infty} \left(\frac{m}{\kappa}\right)^{\beta-1} e^{-(m/\kappa)^{\beta}} \right) \\ & - N(\beta-1)\ln\kappa + (\beta-1)\sum_{i=1}^{N} \ln k_i - \sum_{i=1}^{N} \left(\frac{k_i}{\kappa}\right)^{\beta}. \end{aligned} \tag{A.28}$$

### A.2.4 Normal and Log-Normal Distributions

The normal (or Gaussian) distribution has the form

$$p(k) \sim e^{-\frac{(k-\mu)^2}{2\sigma^2}}, \tag{A.29}$$

where $\mu$ is the mean and $\sigma^2$ is the variance. A similar distribution is the log-normal which is more strongly skewed and is given by

$$p(k) \sim \frac{1}{k} e^{-\frac{(\ln k-\mu)^2}{2\sigma^2}}. \tag{A.30}$$

The log of the likelihood of these are given by

$$\ln \mathcal{L} = N \ln(1-\Delta) - N \ln \left( \sum_{m=k_{\min}}^{\infty} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \right) - \sum_{i=1}^{N} \frac{(k_i-\mu)^2}{2\sigma^2} \tag{A.31}$$

and

$$\ln \mathcal{L} = N \ln(1-\Delta) - N \ln \left( \sum_{m=k_{\min}}^{\infty} \frac{1}{m} e^{-\frac{(\ln m - \mu)^2}{2\sigma^2}} \right) - \sum_{i=1}^{N} \ln k_i - \sum_{i=1}^{N} \frac{(\ln k_i - \mu)^2}{2\sigma^2} \tag{A.32}$$

respectively.

### A.2.5 Poisson Distributions

The Poisson distribution is given by

$$p(k) \sim \frac{\lambda^k}{k!} e^{-\lambda}. \tag{A.33}$$

The log-likelihood is

$$\begin{aligned} \ln \mathcal{L} = N \ln(1-\Delta) - \ln \left( 1 - e^{-\lambda} \sum_{m=0}^{k_{\min}-1} \frac{\lambda^m}{m!} \right) - N\lambda \\ + \ln \lambda \sum_{i=1}^{N} k_i - \sum_{i=1}^{N} \ln(k_i!). \end{aligned} \tag{A.34}$$

In the case that the entire distribution follows a Poisson distribution, then the first and last term vanish and it can be maximised to get

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{A.35}$$

which is just the mean. Otherwise eq. (A.34) has to be numerically maximised to obtain an estimate of $\lambda$.

## A.3 Model Selection

In order to chose the appropriate model to represent the degree distribution, the Akaike and Bayesian Information Criteria are used as described in section 2.3. The standard form of the Akaike Information Criterion is given by

$$\text{AIC} = -2 \ln \mathcal{L}(\hat{\theta}|k_i) + 2n_\theta, \tag{A.36}$$

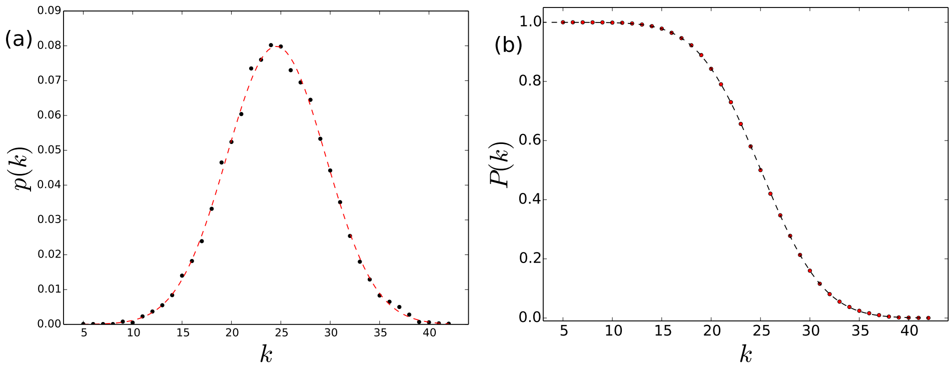



*Fig. A.1:* Panel (a) shows probability distribution for 10,000 samples generated from a normal distribution. The $\mathrm{AIC_c}$ give the strongest support for a normal distribution which is shown as the dashed line. Panel (b) shows the complementary cumulative form of the same data.

where $n_\theta$ is the number of parameters. The version given in section 2.3 however, is is the Akaike Information Criterion with a correction for finite sample sizes, $\mathrm{AIC_c}$. This gives greater penalty for extra parameters. Burnham & Anderson (2002) recommend using the $\mathrm{AIC_c}$ always as if the sample size is large the final term in eq. (2.29) vanishes. For smaller samples the AIC can provide stronger support for models with more parameters than the $\mathrm{AIC_c}$.

Testing the $\mathrm{AIC_c}$ for the models in section 2.2.2, 10,000 points are drawn from a normal distribution with mean $\mu = 25$ and standard deviation $\sigma^2 = 5$. The weights from eq. (2.30) yield $w_{\mathrm{Gaussian}} = 1.0$ indicating that they give full support for the Gaussian/normal distribution. The MLEs provide estimates of $\hat{\mu} = 24.6 \pm 0.1$ and $\hat{\sigma}^2 = 5.0 \pm 0.1$ in close agreement to the actual values. This distribution is shown with a fit generated by the MLEs in fig. A.1 (showing both the probability distribution and the complementary cumulative distribution).

In some cases however there is no model that is completely supported. For example, in the degree distribution for the *Aeneid* shown in fig. 5.16, the $\mathrm{AIC_c}$ weights give support for both a log-normal distribution, $w_{\mathrm{log-norm}} = 0.53$, and a Weibull distribution, $w_{\mathrm{Weibull}} = 0.46$.

Schwarz (1978) suggests that the Bayesian Information Criterion is used instead of the AIC. The BIC penalises parameters more strongly than the AIC. Burnham & Anderson (2002) however argue the $\mathrm{AIC_c}$ has some theoretical advantages over the BIC. As a result, both the $\mathrm{AIC_c}$ and the BIC are used for model selection. However in the data presented here, they always provide support in the same order.

# BIBLIOGRAPHY


Aarne, A., & Thompson, S. (1961). The types of the folktale. 2nd rev. *FF Communications*, *184*.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716–723.

Alberich, R., Miro-Julia, J., & Rossello, F. (2002). Marvel universe looks almost like a real social network. *arXiv preprint cond-mat/0202174*.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, *74*(1), 47.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*(6794), 378–382.

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152.

Antal, T., Krapivsky, P. L., & Redner, S. (2006). Social balance on networks: The dynamics of friendship and enmity. *Physica D*, *224*(1), 130–136.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barber, E. W., & Barber, P. T. (2006). *When they severed earth from sky: how the human mind shapes myth*. Princeton University Press, Princeton.

Barrat, A., & Weigt, M. (2000). On the properties of small-world network models. *The European Physical Journal B*, *13*(3), 547–560.

Bascom, W. R. (1965). The forms of folklore: Prose narratives. *The Journal of American Folklore*, *78*(307), 3–20.

Bayerschmidt, C., & Hollander, L. (1998). *Njal's Saga*. Wordsworth Classics, Hertfordshire.

Bearman, P. S., Moody, J., & Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, *110*(1), 44–91.

Bernard, H. R., Killworth, P. D., Evans, M. J., McCarty, C., & Shelley, G. A. (1988). Studying social relations cross-culturally. *Ethnology*, *27*(2), 155–179.

Bloomfield, M. W. (1970). *The Interpretation of narrative: theory and practice*, vol. 1. Harvard University Press, Harvard.

Boguñá, M., Pastor-Satorras, R., Díaz-Guilera, A., & Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical Review E*, *70*(5), 056122.

Brandes, U., & Erlebach, T. (2005). *Network analysis: methodological foundations*, vol. 3418. Springer.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-model inference: a practical information-theoretic approach*. Springer.

Byock, J. L. (1999). *The Saga of the Volsungs: The Norse Epic of Sigurd the Dragon Slayer*. Penguin, London.

Byock, J. L. (2005). *The Prose Edda: Norse Mythology*. Penguin, London.

Campbell, J. (1949). *The hero with a thousand faces*, vol. 17. Princeton University Press, Princeton.

Campbell, J. (2001). *Thou art that: Transforming religious metaphor*. New World Library, Novato.

Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological review*, *63*(5), 277.

CELT (2004a). The Second Battle of Mag Tuired. `http://www.ucc.ie/celt/published/T300010/index.html`

CELT (2004b). The Second Battle of Moytura. `http://www.ucc.ie/celt/published/T300011/index.html`

Choi, Y.-M., & Kim, H.-J. (2007). A directed network of Greek and Roman mythology. *Physica A*, *382*(2), 665–671.

Clauset, A. (2007). [Accessed: 2013-06-12]. `http://tuvalu.santafe.edu/~aaronc/powerlaws/`

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661–703.

Cohen, O., Barocas, A., & Geffen, E. (2013). Conflicting management policies for the Arabian wolf Canis lupus arabs in the negev desert: is this justified? *Oryx*, *47*(02), 228–236.

Cohen, R., Erez, K., Ben-Avraham, D., & Havlin, S. (2000). Resilience of the internet to random breakdowns. *Physical review letters*, *85*(21), 4626.

Costa, L. d. F., Oliveira Jr, O. N., Travieso, G., Rodrigues, F. A., Villas Boas, P. R., Antiqueira, L., Viana, M. P., & Correa Rocha, L. E. (2011). Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, *60*(3), 329–412.

Croft, D. P., James, R., & Krause, J. (2008). *Exploring animal social networks*. Princeton University Press, Princeton.

Crossley-Holland, K. (1999). *Beowulf*. Oxford University Press, Oxford.

Davies, S. (2007). *The Mabinogion*. Oxford University Press, Oxford.

Davis, G. F., Yoo, M., & Baker, W. E. (2003). The small world of the american corporate elite, 1982-2001. *Strategic organization*, *1*(3), 301–326.

De Troyes, C. (1991). *Arthurian romances*. Penguin, London.

D'Huy, J. (2013). A phylogenetic approach of mythology and its archaeological consequences. *Rock Art Research*, *30*(1), 115–118.

Dooley, A., & Roe, H. (1999). *Tales of the Elders of Ireland (translation of Acallam na Senórach (c. 1200 AD), with introduction and notes)*. Oxford University Press, Oxford.

Dunbar, R., & Dunbar, P. (1974). Social dynamics of gelada baboons. *Contributions to primatology*, *6*, 1–157.

Dundes, A. (1984). *Sacred narrative: readings in the theory of myth*. University of California Press, Oakland.

Dundes, A. (1997). Binary opposition in myth: The propp/levi-strauss debate in retrospect. *Western folklore*, *56*(1), 39–50.

Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev, V., Buldyrev, S. V., da Luz, M. G., Raposo, E. P., Stanley, H. E., *et al.* (2007). Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, *449*(7165), 1044–1048.

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, (pp. 1429–1451).

Efron, B. (1979). Computers and the theory of statistics: thinking the unthinkable. *Siam Review*, *21*(4), 460–480.

Eliade, M. (1998). *Myth and reality*. Waveland Press, Long Grove.

Elson, D. K., Dames, N., & McKeown, K. R. (2010). Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (pp. 138–147). Association for Computational Linguistics.

Erdős, P., & Rényi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, *6*, 290–297.

Estrada, E. (2011). *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, Oxford.

E.V. Rieu (2003). *The Iliad*. Penguin Classics, London.

Fararo, T. J., & Sunshine, M. H. (1964). *A study of a biased friendship net*. Syracuse University Youth Development Center, New York.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*, 309–368.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3), 75–174.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, (pp. 35–41).

Fronczak, A., Fronczak, P., & Hołyst, J. A. (2004). Average path length in random networks. *Physical Review E*, *70*(5), 056110.

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M., & Souma, W. (2004). Do Pareto–Zipf and Gibrat laws hold true? An analysis with European firms. *Physica A*, *335*(1), 197–216.

Galam, S. (2008). Sociophysics: A review of galam models. *International Journal of Modern Physics C*, *19*(03), 409–440.

Galam, S., Gefen, Y., & Shapir, Y. (1982). Sociophysics: A new approach of sociological collective behaviour. I. Mean-behaviour description of a strike. *Journal of Mathematical Sociology*, *9*(1), 1–13.

Gantz, J. (1981). *Early Irish Myths and Sagas*. Penguin, London.

George, A. (2002). *The Epic of Gilgamesh*. Penguin, London.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826.

Gleiser, P. M. (2007). How to become a superhero. *Journal of Statistical Mechanics: Theory and Experiment*, *2007*(09), P09020.

Gleiser, P. M., & Danon, L. (2003). Community structure in jazz. *Advances in complex systems*, *6*(04), 565–573.

Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical review E*, *68*(6), 065103.

Hamer, A. J. (2008). *Njáls saga and its Christian background: a study of narrative method*. University of Groningen, Groningen.

Hatto, A. (2004). *The Nibelungenlied*, vol. 137. Penguin, London.

Hazelkorn, E., Ryan, M., Gibson, A., & Ward, E. (2013). Recognising the value of the arts and humanities in a time of austerity.

Heaney, S. (1999). *Beowulf*. Faber & Faber, London.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology*, *21*(1), 107–112.

Humphries, M. D., & Gurney, K. (2008). Network small-world-ness: a quantitative method for determining canonical network equivalence. *PLoS One*, *3*(4), e0002051.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481).

Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., & Van den Broeck, W. (2011). What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, *271*(1), 166–180.

Jeong, H., Mason, S. P., Barabási, A.-L., & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42.

Kaiser, M. (2008). Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. *New Journal of Physics*, *10*(8), 083042.

Kinsella, T. (1969). *The Táin*. Oxford University Press, Oxford.

Kirk, G.S. (1980). *The Odyssey*. Oxford University Press, Oxford.

Klimt, B., & Yang, Y. (2004). The Enron corpus: A new dataset for email classification research. In *Machine learning: ECML 2004*, (pp. 217–226). Springer.

Knappett, C., Evans, T., & Rivers, R. (2008). Modelling maritime interaction in the aegean bronze age. *Antiquity*, *82*(318).

Knuth, D. E., Knuth, D. E., & Knuth, D. E. (1993). *The Stanford GraphBase: a platform for combinatorial computing*, vol. 4. Addison-Wesley Reading.

KONECT (2013). Slashdot Zoo network dataset. [Accessed: 2014-02-02]. `http://konect.uni-koblenz.de/networks/slashdot-zoo`

Kousoulis, A. A., Economopoulos, K. P., Poulakou-Rebelakou, E., Androutsos, G., & Tsiodras, S. (2012). The plague of Thebes, a historical epidemic in Sophocles Oedipus Rex. *Emerging infectious diseases*, *18*(1), 153.

Kumar, R., Novak, J., & Tomkins, A. (2010). Structure and evolution of online social networks. In *Link Mining: Models, Algorithms, and Applications*, (pp. 337–357). Springer.

Kunegis, J., Lommatzsch, A., & Bauckhage, C. (2009). The slashdot zoo: mining a social network with negative edges. In *Proceedings of the 18th international conference on World wide web*, (pp. 741–750). ACM.

Kuperman, M., & Abramson, G. (2001). Small world effect in an epidemiological model. *Physical Review Letters*, *86*(13), 2909.

Larrington, C. (1999). *The Poetic Edda*. Oxford University Press, Oxford.

Latapy, M. (2008). Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science*, *407*(1), 458–473.

Leeming, D. (2005). *The Oxford companion to world mythology*. Oxford University Press, Oxford.

Leeming, D. A. (1991). *The World of Myth: An Anthology*. Oxford University Press, Oxford.

Leskovec, J., Lang, K. J., Dasgupta, A., & Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, *6*(1), 29–123.

Lévi-Strauss, C. (1955). The structural study of myth. *Journal of American folklore*, (pp. 428–444).

Lindow, J. (2002). *Norse Mythology: A Guide to Gods, Heroes, Rituals, and Beliefs*. Oxford University Press, Oxford.

Luce, R. D., & Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika*, *14*(2), 95–116.

Luduena, G. A., Meixner, H., Kaczor, G., & Gros, C. (2013). A large-scale study of the world wide web: network correlation functions with scale-invariant boundaries. *European Physical Journal B*, *86*, 348.

Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(Suppl 2), S186–S188.

Lyle, E. (2006). Narrative form and the structure of myth. *Folklore: Electronic Journal of Folklore*, (33), 59–70.

Lyle, E. (2012). *Ten Gods: A New Approach to Defining the Mythological Structures of the Indo-Europeans*. Cambridge Scholars Publishing, Cambridge.

Mac Carron, P., & Kenna, R. (2012). Universal properties of mythological networks. *EPL (Europhysics Letters)*, *99*(2), 28002.

Mac Carron, P., & Kenna, R. (2013a). Network analysis of the Íslendinga sögur–the sagas of Icelanders. *The European Physical Journal B*, *86*(10), 407.

Mac Carron, P., & Kenna, R. (2013b). A quantitative approach to comparative mythology. *Cosmos*, *14*, 103–117.

Mac Carron, P., & Kenna, R. (2013c). Viking sagas: Six degrees of Icelandic separation. social networks from the viking era. *Significance*, *10*(6), 12–17.

Mac Carron, P., Platini, T., & Kenna, R. (2014). Paper in preparation.

MacKillop, J. (2004). *A dictionary of Celtic mythology*. Oxford University Press, Oxford.

Mahnusson, M., & Pálsson, H. (1960). *Njal's Saga*. Penguin, London.

Mansfield, E. (1962). Entry, Gibrat's law, innovation, and the growth of firms. *The American Economic Review*, *52*(5), 1023–1051.

Matarasso, P. M. (1969). *The Quest of the Holy Grail*. Penguin, London.

McAuley, J., & Leskovec, J. (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems 25*, (pp. 548–556).

McCafferty, P., & Baillie, M. (2005). *The Celtic gods: comets in Irish mythology*. Tempus Publishing.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, (pp. 415–444).

Meyer, K. (1910). *Fianaigecht: being a collection of hitherto inedited Irish poems and tales relating to Finn and his Fiana, with an English translation*, vol. 16. Hodges, Figgis.

Milgram, S. (1967). The small world problem. *Psychology today*, *2*(1), 60–67.

Miranda, P. J., Baptista, M. S., & Pinto, S. E. d. S. (2013). Analysis of communities in a mythological social network. *arXiv preprint arXiv:1306.2537*.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, (pp. 29–42). ACM.

Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random structures & algorithms*, *6*(2-3), 161–180.

Moreno, J. L. (1934). *Who shall survive?: A new approach to the problem of human interrelations.*. Nervous and Mental Disease Publishing Co.

Morford, M. P., & Lenardon, R. J. (1999). *Classical mythology*. Oxford University Press, Oxford.

Müller, F. M. (1866). *Lectures on the science of language: delivered at the Royal Institution of Great Britain in April, May, & June 1861*. Longmans, Green.

Murphy, G. (1961). *Saga and myth in ancient Ireland*, vol. 10. Published for the Cultural Relations Committee of Ireland by Colm O Lochlainn.

Newman, M. (2011). Complex systems: A survey. *Am. J. Phys.*, *79*, 800–810.

Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E*, *64*(1), 016131.

Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, *74*(3), 036104.

Newman, M. E. J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409.

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical review letters*, *89*(20), 208701.

Newman, M. E. J. (2003a). Mixing patterns in networks. *Physical Review E*, *67*(2), 026126.

Newman, M. E. J. (2003b). The structure and function of complex networks. *SIAM review*, *45*(2), 167–256.

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary physics*, *46*(5), 323–351.

Newman, M. E. J. (2009). *Networks: An Introduction*. Oxford University Press, Oxford.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, *69*(2), 026113.

Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, *68*(3), 036122.

Nilsson, M. P. (1932). *The Mycenaean origin of Greek mythology*. University of California Press, Oakland.

Oakley, M. (2004). *The Aeneid*. Wordsworth Classics, Hertfordshire.

O'Bryan, A., *et al.* (1993). *Navaho Indian Myths*, vol. 163. Courier Dover Publications, New York.

O'Donoghue, H. (2004). *Old Norse-Icelandic Literature: A Short Introduction*. Blackwell Publishing, Oxford.

O'Rahilly, T. F. (1946). *Early Irish history and mythology*. Dublin Institute for Advanced Studies, Dublin.

Palchykov, V., Kertész, J., Dunbar, R., & Kaski, K. (2013). Close relationships: A study of mobile communication records. *Journal of Statistical Physics*, *151*, 735–744.

Pálsson, H., & Edwards, P. G. (1981). *Orkneyinga saga: the history of the Earls of Orkney*. Penguin, London.

Ravasz, E., & Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical Review E*, *67*(2), 026112.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., *et al.* (1997). Protecting adolescents from harm. *JAMA: the journal of the American Medical Association*, *278*(10), 823–832.

Rocha, L. E., Liljeros, F., & Holme, P. (2011). Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS computational biology*, *7*(3), e1001109.

Schnettler, S. (2009). A structured overview of 50 years of small-world research. *Social Networks*, *31*(3), 165–178.

Schrempp, G., & Hansen, W. (2002). *Myth: A New Symposium*. Indiana University Press, Bloomington.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

Smiley, J. (2000). *The Sagas of Icelanders: A Selection*. Penguin, London.

Stiller, J., Nettle, D., & Dunbar, R. I. (2003). The small world of shakespeares plays. *Human Nature*, *14*(4), 397–408.

Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*(6825), 268–276.

Stumpf, M. P., & Porter, M. A. (2012). Critical truths about power laws. *Science*, *335*(6069), 665–666.

Šubelj, L., & Bajec, M. (2012). Clustering assortativity, communities and functional modules in real-world networks. *arXiv preprint arXiv:1202.3188*.

Szell, M., & Thurner, S. (2010). Measuring social dynamics in a massive multiplayer online game. *Social networks*, *32*(4), 313–329.

Sznajd-Weron, K. (2005). Sznajd model and its applications. *Acta Physica Polonica B*, *36*(8).

Tehrani, J. J. (2013). The phylogeny of Little Red Riding Hood. *PloS one*, *8*(11), e78871.

Tolkien, J. R. R. (1947). *On fairy-stories*. Oxford University Press, Oxford.

Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). On the evolution of user interaction in Facebook. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks (WOSN'09)*.

von Strassburg, G. (2004). *Tristan with the 'Tristan' of Thomas*. Penguin, London.

Wasserman, S. (1994). *Social network analysis: Methods and applications*, vol. 8. Cambridge university press, Cambridge.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, *393*(6684), 440–442.

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *314*(1165), 1–340.

Witzel, M. (2012). *The Origins of the World's Mythologies*. Oxford University Press, Oxford.