Comparing social networks of translations: a case study on the medieval irish epic Táin Bó Cúailnge

Social Network Analysis and Mining (2026) 16:21
https://doi.org/10.1007/s13278-025-01547-3

ORIGINAL ARTICLE

Comparing social networks of translations: a case study on the medieval irish epic *Táin Bó Cúailnge*

Stephen McKermitt[1] · Gaëlle Clion[2,3] · Pádraig MacCarron[1]

Received: 12 August 2025 / Revised: 19 October 2025 / Accepted: 23 October 2025

Abstract

This work presents a comparative analysis of the characteristics of social networks derived from different versions and translations of the Irish epic *Táin Bó Cúailnge*. The characteristics of the social networks are compared using graph-matching techniques and comparing structural properties, as well as truncating the larger networks to match the size of the smaller networks for further comparison. We find that the original Irish edition and the direct English translation of the *Táin Bó Cúailnge* Recension I by Cecile O'Rahilly are most similar to one another, and that the modern English versions have the furthest distance from the sources. Individually we show that each social network, regardless of the author or network size, is disassortative and display "small-world" network characteristics. They are also shown to have largely similar degree distributions, with Maximum Likelihood Estimate fitting showing that most networks can be fitted using power law distributions. The results show that the more similar a version stays to the source material, the more similar the generated social networks will be.

Keywords Social networks · Complex networks · Narrative networks · Graph matching · Celtic studies

1 Introduction

The study of narrative networks, or character networks, has become increasingly common in recent years (Labatut and Bost 2019). These studies involve many approaches, for example, using natural-language processing to extract social networks to compare texts (Janickyj 2023), assessing literary theories (Elson et al. 2010), to comparing the structural-network properties of narratives (Mac Carron and Kenna 2012). However, while extracting the data is a major issue in these studies, different translations or interpretations of the data source are rarely compared to assess how robust the network extraction is.

In this study, we compare two medieval versions of an Irish narrative text recorded a few centuries apart, as well as four English translations of that same epic. In each case, we compare the social networks extracted from them to ascertain if there are any substantial differences, and whether previous work is safe to rely on translations, or instead should stick to the source where possible. To do these comparisons, we compare standard complex network properties (similar to Refs. Massey (2019); Mac Carron (2014)) as well as using *graph matching* techniques to quantify how different their social networks are.

Graph matching is a relatively old problem, predating the concept of *complex networks* by a few decades. There are many different methods for graph matching, for example tree searches (eg. Sanfeliu and King-Sun (1983)), spectral methods (eg. Ipsen (2004)), and more recently methods using mutual information (eg. Felippe et al. (2024)). For a detailed —but now somewhat old— survey on different approaches to graph matching algorithms, see Ref. Conte

Gaëlle Clion and Pádraig MacCarron have contributed equally to this work.

Stephen McKermitt
mckermitt.stephen@ul.ie

Gaëlle Clion
gaelle.clion@ul.ie

Pádraig MacCarron
padraig.maccarron@ul.ie

1 MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

2 School of Celtic Studies, Dublin Institute for Advanced Studies, Dublin, Ireland

3 School of English, Irish, and Communication, University of Limerick, Limerick, Ireland

Published online: 24 December 2025

(2004). These methods all quantify the similarity of the graphs with a single value. Hence, as well as using some of these methods, we also look at some standard structural network properties (such as the degree distributions, small worldness, and assortativity) to further compare the networks herein.

Similar approaches can be found in Ref. Amalvy (2024) where the same stories in different media are compared. There, the characters networks of George R.R. Martin's *A Song of Ice and Fire*, the comic series of the same name, and the TV show *Game of Thrones* are compared. In that case, textual information (such as affiliation) was required to match characters in the different media. Traditional network methods of comparing differences in texts often involve using semantic networks, see for example Refs (Danowski 1993; Carley 1993; Doerfel and Barnett 1999). Kwon et al. Kwon et al. (2009) use both semantic networks and spatial analysis to show subtle differences in interpretation between translations of the universal declaration of human rights. In the narrative of interest here however, the medieval versions are not translations of one another, but different retellings of the same story. We are interested in whether the social networks differ, rather than the events or language. As this narrative fits into a larger corpus of texts with a similar cast of characters, often referred to as *The Ulster Cycle*, we are also interested in how the version we use would affect a larger network as in Ref. MacCarron (2024).

Another related work is the comparison of the social networks of the four gospels (Massey 2019). There it was found that the structural properties of the gospel according to John is (unsurprisingly) different from the other three. That study also relied on English translations and one particular version of each. Recently, a similar idea to ours compared the character networks of *Don Quixote*, to both a "youth" version and a children's version (Lei et al. 2025). Like in Massey's aforementioned gospel comparison (Massey 2019), they just focus on structural properties rather than also using graph-matching techniques to quantify a difference.

Here, we are interested in comparing the social networks of medieval Irish narratives. These narratives are mostly preserved in manuscripts, transcribed by monks and scribes during the medieval and early modern period. One of the most well-known of these stories is the epic *Táin Bó Cúailnge* (*The Cattle Raid of Cooley*, often abbreviated to *The Táin*). Likely first compiled in the ninth century, this story portrays Queen Medb of Connacht's attempt to steal the prized Brown Bull of Cooley, leading to a great war against the men of Ulster. The hero, Cú Chulainn, single-handedly defends Ulster through a series of battles, showcasing his super-human strength while the men of Ulster are under a curse which leaves them unable to defend themselves (Wright 2011).

The Táin appears in at least ten different manuscripts — sometimes only in fragments— which can be divided into three different groups showcasing different recensions of the tale. Various editions and translations of *The Táin* from Irish to English have been completed, some including related stories, known as *remscéla* or pre-tales. This study uses a number of these translations, as well as editions of the text in the original Irish. Cecile O'Rahilly's editions of Recensions I and II provide a baseline for what the original text of *The Táin* looks like (O'Rahilly 2001a, b, 2000). Thomas Kinsella (Kinsella 1970) and Ciaran Carson's (Carson 2007) versions of *The Táin* are modern translations of the different episodes found in Recensions I and II. Joseph Dunn's English translation (Dunn 1914) differs from these two, as it fully combines readings from all recensions, including Recension III. O'Rahilly's translation used here is a direct translation of Recension I based on her own edition of the text. Carson's translation is primarily based on O'Rahilly's editions of Recension I, as well as elements from Recension II; while Kinsella makes use of her edition of Recension II, but follows an older edition of Recension I, that of John Strachan and J.G. O'Keeffe (Strachan and O'Keeffe 1912), as O'Rahilly's one was not published yet. Kinsella also includes a selection of pre-tales from other narratives in that literary tradition that lead up to the events of *the Táin* in order to provide context for these events. Dunn on the other hand is an older translation, following a different methodology, and making use of a variety of sources, such as editions from Ernst Windisch and Max Nettlau, as well as Strachan and O'Keeffe mentioned above. He seems to have also consulted some of the manuscripts themselves. His approach is most notable for his complete combination of the different versions of the story throughout the text, rather than selecting passages from one recension or the other. This translation showcases a method for translating medieval texts with a much more important involvement and impact of the translator on the text.

In order to compare them, this study investigates the social networks generated using data from the different translations and editions of *The Táin*, focusing on quantifying how similar each of the networks are using different metrics. We begin by using graph-matching techniques as discussed above as well as analysing the different structural properties of these editions and translations. Kinsella's translation of *The Táin* has been used in previous works (Mac Carron and Kenna 2012). However, it has not previously been compared with other translations or editions to determine whether it is a reasonable version to use. The findings of this study will provide a broader understanding of how translation influences social network structures.

2 Methods

To gather the networks, the data were extracted manually. Each interaction was recorded in a dataset, noting the characters involved and the nature of their relationship (e.g., friendly, hostile). A social network was generated for each version, where *nodes* represent characters and the *edges* represent interaction. We treat "hostile" interactions (i.e. when two characters fight in combat) as negative and all others as positive. We weight edges based on how many chapters they co-occur. However, we do not use weights or sign in this particular analysis.

Characters are frequently referenced using alternate names and slight variations in spellings of the same name throughout the texts. In order to account for this, the alternate names were noted during data collection to avoid duplication.

We use the following editions and translations in this analysis: the online version of O'Rahilly Recension I (abbreviated to ORI) (O'Rahilly 2001a) which is in Irish, the online version of O'Rahilly's English translation of this (ORI-Eng) (O'Rahilly 2001b), the online version of O'Rahilly's Recension II (ORII) (O'Rahilly 2000), the combined network of these (OR-comb), Kinsella's version (Kin) (Kinsella 1970), Kinsella's version without pretales (Kins-X), Carson's version (Cars) (Carson 2007) and Dunn's version (Dunn) (Dunn 1914). These abbreviations can be found in Table 5 in the Appendix for the reader's convenience.

2.1 Network comparison

The networks are compared in two different ways. We begin by reporting the structural network properties to ascertain where the social networks are similar and differ from one another. We then use different graph-matching techniques to quantify the difference in the overall networks. We then return to a more traditional degree-distribution analysis as well as analyse the centrality of some protagonists in each version. Finally, we perform some simulations by reducing the networks to the same size to see if they become more similar.

For a network with N nodes and L edges, the average degree $\langle k \rangle$ represents the average number of edges connected to each node in the network. The second moment $\langle k^2 \rangle = 1/N \sum_{i=1}^{N} k_i^2$ is related the variance of the network $(\langle k^2 \rangle - \langle k \rangle^2)$. When plotting the degree distribution of a network with a high degree variance, it tends to skew right. This skew can be seen in all networks presented here, as *The Táin* centres around a small number of main characters who are connected to a plethora of minor characters. Hence, we expect a high value of $\langle k^2 \rangle$.

The *assortativity* of a network, r, is the degree-degree correlations at the end of an edge and defines the likelihood that a node connects to other nodes with similar degrees (Newman 2002). Each edge contributes positively to the assortativity if the degree of the nodes at either edge are both below or above $\langle k^2 \rangle / \langle k \rangle$ (which is the average degree at the end of an edge) (MacCarron et al. 2023). If the degrees of the pair of nodes at the end of an edge are different sides of this average, then that edge contributes disassortatively. A network is assortative if $r > 0$ and disassortative if $r < 0$ (indicating the nodes tend to connect to nodes with non-similar degrees).

The *mean clustering coefficient,* $\langle C \rangle$ quantifies how much nodes cluster together in a network. This is the mean of the local clustering coefficients (Watts and Strogatz 1998). Social networks in particular tend to have high clustering coefficients (Albert and Barabási 2002). The *average path length*, ℓ, of a network is the mean geodesic between all pairs of nodes (Newman 2013).

Social networks have a tendency to display *small-world* characteristics and are often *assortative* (Newman and Park 2003). A network being 'small-world' implies that it has a much larger clustering coefficient compared to a random network of the same size and has an average path length that is approximately equal to that of a random Erdős-Rényi graph (Watts and Strogatz 1998). The exact form of the random mean path length ℓ_{rand} is given in Ref. (Fronczak et al. 2004) and C_{rand} in Ref. Albert and Barabási (2002).

As mentioned earlier, there are many approaches to graph matching. Although we are working with the same narrative, no two versions contain the same set of characters, hence we are interested in inexact matching (Conte 2004). We first compute the *graph edit distance* which calculates the minimum number of changes needed to transform one network into the other (Sanfeliu and King-Sun 1983). A lower edit distance indicates a greater magnitude of similarity, as fewer structural changes are necessary to align the two networks. We normalise this by the maximum graph edit distance (i.e. divide the value by the number of edges in the complete graph or the larger graph added to the difference in the number of nodes of the two graphs). We report the actual graph edit distance in the appendix.

We next compute the Ipsen-Mikhailov (IM) distances for each pair of networks. IM distance is a spectral method that computes the Lorenz distance between two networks as a means of quantifying their differences (i.e., it compares the networks by computing the integral of the difference of the Lorentz function (Ipsen 2004)). This computes a value between 0 and 1, where 0 indicates identical network structures, and 1 represents maximal dissimilarity, such as when one network is an empty graph (ie, no edges) and the other is a graph with $N(N-1)/2$ edges. In summary, a smaller IM distance indicates closer structural similarity between networks. IM distances have been used in previous literature to compare different types of networks (Donnat and Holmes 2018).

Finally for the graph matching, we use the recently developed network mutual information to determine graph similarity (Felippe et al. 2024). It should be noted that this method requires node-aligned graphs. However, as there is a significant overlap among these, we *pad* the networks by adding the missing nodes with no edges to calculate the mutual information. The network mutual information is 1 when the two networks are identical, hence we report one minus the network mutual information to have a value comparable to the normalised graph edit distance and the IM distance.

We then return to some standard structural network analysis. The *degree distribution* of a network represents the probability for a node to have degree k. Maximum likelihood estimation was used to identify which distribution best fits the data from a selection of predefined distributions (Mannion and MacCarron 2023; Clauset et al. 2009) as a method of comparing the different networks. Using this method allows conclusions to be made about whether the different networks can be fitted to similar degree distributions.

Two-sample Kolmogorov-Smirnov (KS) tests are also used to compute a distance measure between two distributions as yet another method of comparison. The null hypothesis for the KS test is that the two datasets are drawn from the same distribution. When a small p-value is computed, this indicates that it is unlikely that they are from the same distribution. More importantly for us here, the KS statistic provides a distance between two distributions. This approach complements the maximum likelihood estimation to fit the most appropriate distribution to data.

On a node-centric level, the characters with the highest degree and betweenness for each of the networks are compared. The betweenness centrality of a node is the number of shortest paths passing through that node (Newman 2003). Within the context of this paper, it can be used to find a character's influence within the rest of the network. To compare these lists of characters, we use Jaccard coefficient to see how similar the sets of the top ten characters are. We then use the Spearman's rank coefficient to determine if the order is consistent between both.

Finally, to further compare networks of differing sizes, a random sampling procedure was performed to observe how structural properties change with network size. The larger networks were down-sampled by randomly removing nodes until their size matches that of the smaller network. This random removal process is repeated 100 times, after which the network properties were averaged to give an estimate of the properties of the size-matched network. This allows for a comparison to be made between the networks if they were the same size.

3 Results

Table 1 shows the network properties found in the networks used. During the data gathering process it was observed that each of the editions and translations contain lengthy lists of side characters mentioned only by name during the section of the text which involves the mustering of the armies of Connacht and Ulster. Many of these characters are mentioned only once, some share identical names; "Ailill mac Máta chanted these words: 'Arise, O Traigthrén! I send you to the three Conaires from Slíab Mis, the three Lesfinds from Lúachair; the three Meid Corpthe Loste, the three..." O'Rahilly (2001b). Their presence does not contribute to the broader social structure of the tale. For this reason, it was decided that these lists of side characters would be omitted from the network analyses which are shown in the bottom part of Table 1.

These lists are much more prominent in the original recensions (ORI and ORII) compared to the modern translations (Kin, Cars and Dunn), they also have a considerable effect on quantities such as the assortativity r due to the removal of many low-degree characters who are only connected to their ruler, who are far on the right of the degree distribution.

Kinsella's relatively large number of edges can be attributed to the inclusion of the pre-tales which focuses on introducing peripheral characters and exploring their backgrounds. The pre-tales introduce more interactions between characters, increasing the number of edges in the network compared to the other versions. When the characters from the pre-tales are removed from Kinsella's network, the number of edges is more similar to that of the O'Rahilly networks. This can be explained by the main text of Kinsella's version using O'Rahilly as one of his sources. With that being said, there is a notable difference in the number of edges when comparing Kins, Kins-X to the O'Rahilly networks. Even without the pre-tales, Kinsella's version has a much larger number of edges compared to all of the individual O'Rahilly networks. This could potentially be attributed to the writing style of Kinsella. In an attempt to make a more cohesive story, Kinsella may have introduced additional connections. It could also stem from Kinsella using a different edition than O'Rahilly's for Recension I. O'Rahilly, on the other hand, is a direct edition of the original texts. When recensions I and II are combined, the resulting network has characteristics that are much closer to the networks of Kinsella than when the recensions are separated. Carson's network on the other hand is more similar to O'Rahilly's Recensions

Table 1 Demonstrates the global properties of the generated social networks

Property	N	L	$\langle k \rangle$	$\langle k^2 \rangle$	r	$\langle C \rangle$	C_{rand}	$\langle \ell \rangle$	ℓ_{rand}	γ
Original										
ORI-Eng	505	999	3.96	212.63	-0.51	0.45	0.01	2.73	4.61	2.22
ORI	507	1000	*3.94*	214.94	*-0.52*	*0.44*	0.01	2.73	**4.62**	2.23
ORII	409	1099	5.37	225.87	-0.46	0.61	0.01	*2.7*	3.73	2.22
OR-comb	**654**	**1573**	4.81	**319.95**	-0.5	0.53	0.01	*2.7*	4.26	*2.17*
Kins	422	1266	**6.00**	207.76	**-0.35**	0.73	0.01	2.8	3.55	2.28
Kins-X	*354*	1025	5.79	198.45	-0.42	**0.76**	**0.02**	*2.7*	*3.51*	**2.35**
Cars	441	*954*	4.33	*192.44*	-0.51	0.66	0.01	**2.83**	4.26	2.3
Dunn	537	1425	5.31	291.80	-0.43	0.61	0.01	2.64	3.92	2.19
No Lists										
ORI-Eng	255	594	4.66	128.73	-0.31	0.43	0.02	2.64	3.73	2.28
ORI	257	*593*	*4.61*	131.63	-0.31	*0.42*	0.02	*2.63*	**3.75**	2.30
ORII	250	723	5.78	*113.67*	**-0.29**	0.49	0.02	2.80	3.32	*2.02*
OR-comb	353	1001	5.67	**189.04**	-0.3	0.47	0.02	2.68	3.55	2.03
Kins	**395**	**1053**	5.33	178.06	-0.32	0.49	*0.01*	2.79	3.73	2.13
Kins-X	*235*	710	**6.04**	140.34	-0.32	**0.65**	**0.03**	2.65	*3.21*	2.18
Cars	267	625	4.68	125.18	*-0.36*	0.57	0.02	**3.08**	3.75	**2.40**
Dunn	302	837	5.54	154.88	-0.32	0.54	0.02	2.68	3.50	2.16

The datasets originally contain long lists of side characters, the first table includes these characters, the second table shows the values when the groups are combined into a single node. The column with γ shows the exponent of a fitted power law. The largest values in the columns for each version of the datasets are in **bold**, the smallest values are *italicised*

I and II in scale, containing a similar number of nodes and edges to the individual networks. This can be explained by Carson compiling pieces from O'Rahilly's Recensions I and II to create his own version of the tale. Dunn's edition of *The Táin* takes elements of all three recensions, and has the greatest number of nodes and edges after Kinsella's network and the combination of the two O'Rahilly recensions.

Despite their differences, all networks are shown to be disassortative, which is not common for social networks (Newman and Park 2003), but common in these ancient narratives (Mac Carron 2014). Additionally, all networks have similar values for their assortativity coefficients.

All the networks have similar values for average path lengths and clustering coefficients. They display small-world properties (i.e., when compared to a random network of the same size, their clustering coefficient is larger, and their average path length is smaller). It is worth expanding on the mean path length $\langle \ell \rangle$ being shorter than the random Erdős-Rényi value. As discussed later, power laws are found to be the most likely of a few candidate models for the degree distribution for the majority of these datasets. Fronczak *et al.* Fronczak et al. (2004) estimates the expected mean path length for the Barabási-Albért model ℓ_{BA} in eq. (19). Testing that for ORI for example, we get $\ell_{\text{BA}} = 2.75$ which is closer to the observed value of $\langle \ell \rangle = 2.63$ than $\ell_{\text{rand}} = 3.75$. Many power-law degree distributed networks are said to be *ultra-small world* (Cohen and Havlin 2003). Traditionally, this computes the diameter rather than the mean path length (eg. Cohen and Havlin (2003)). Recently, van der Hofstad1 & Komjáthy. van der Hofstad and Komjáthy (2017) estimate the typical mean path length for an ultra-small power-law network to be $\hat{\ell} = 2 \log \log N / |\log(\gamma - 2)|$. For ORI, this gives $\hat{\ell} = 1.46$, indicating that these networks are not ultra-small world.

The γ values highlight that when the degree distributions of each network are forced to be power laws, their coefficient values are all similar.

3.1 Graph matching

To assess how similar the networks are, we begin by using the normalised graph edit distance. We next use the IM distances, a spectral method to compare similarities between the networks. Finally we compute the network mutual information (NMI) values (and show 1-NMI to keep them all within the same range for interpretation). These were adjusted such that for each of the metrics, lower values indicate that less edges need to be edited to make the graphs exact. Table 3.1 presents the results of the network comparisons.

For each of the metrics, we see that again the English and Irish O'Rahilly Recension I are the most similar. The Kinsella networks (both with and without the pre-tales) are the least similar. Generally, when comparing each of the O'Rahilly networks to one another, they are closer than comparing the O'Rahilly networks to Carson's, Kinsella's and Dunn's networks.

The table of IM values show that these distances across all pairs are below 0.2, indicating the networks share similar structures without extreme dissimilarities present. IM distance is 1 when comparing an empty graph to a complete graph, so the values will always be low for sparse networks (Table 2).

3.2 Degree distributions

Maximum likelihood estimates are used to identify which distribution out of a predefined list best fits each of the networks. These distributions include discrete versions of the power law, exponential, Weibull, normal, Poisson, stretched exponential, truncated-power law (i.e. a power law with an exponential cut-off) and lognormal. The best candidate distribution are primarily power-law distributions. O'Rahilly's Irish Recension II network is the only exception being fitted by a Weibull distribution instead. Figure 1 shows the and complementary cumulative distribution functions for O'Rahilly, Kinsella, Carson and Dunn's networks, where the datasets including the large groups of characters are on the left, and without are on the right. They have a large tail on the right hand side for all datasets. The scatter plot shows that the probability distributions for the networks are similar. However, there is a distinct set of outliers in Dunn and O'Rahilly Recension II's networks just under $k = 20$, which is not present for the Kinsella or O'Rahilly English Recension I plots, which manifests itself as larger values in the KS test results (Table 3). The probability mass functions for all of the datasets (including O'Rahilly's combined network, Kinsella without pretales, O'Rahilly's English translation of Recension I) can be seen in the Appendix (Fig. 2) and here it is observed there are numerous characters with the same degree of 17 resulting in the outliers.

When forcing all networks to match a power law by setting $k_{min} = 2$, their exponent values (γ) are all similar (≈ 2), as seen in Table 1. It is easy to show that a continuous power law with $\gamma < 2$ has an infinite first moment (i.e. mean degree) (Newman 2005) (while with $2 < \gamma < 3$, the mean is finite, but the second moment infinite). As the networks presented here are finite (all of the order of a few hundred nodes), as a result the continuous approximation may not be a reliable model as can be seen from the small values of mean degrees in Table 1.

As well as fitting the distributions, we also use the KS statistic to obtain a distance between each degree distribution.

Table 2 All of the values for the graph matching section of the results

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	CARS	Dunn
Normalised GED								
ORI-Eng	**0**							
ORI	**0.01**	**0**						
ORII	0.03	0.03	**0**					
OR-comb	0.02	0.02	0.01	**0**				
Kins	0.02	0.02	0.02	0.02	**0**			
Kins-X	0.03	0.03	*0.04*	0.02	0.02	**0**		
Cars	0.03	0.03	0.03	0.02	0.02	0.03	**0**	
Dunn	0.03	0.02	0.03	0.02	0.02	0.03	0.02	**0**
IM Distance								
ORI-Eng	**0**							
ORI	**0.01**	**0**						
ORII	0.06	0.07	**0**					
OR-comb	0.04	0.05	0.02	**0**				
Kins	0.04	0.04	0.04	0.03	**0**			
Kins-X	*0.1*	*0.1*	0.08	0.08	0.07	**0**		
Cars	0.04	0.04	0.07	0.06	0.05	0.08	**0**	
Dunn	0.04	0.04	0.04	0.03	0.04	0.08	0.04	**0**
1-NMI								
ORI-Eng	**0**							
ORI	**0.12**	**0**						
ORII	0.74	0.74	**0**					
OR-comb	0.42	0.37	0.39	**0**				
Kins	0.99	0.99	*0.99*	0.99	**0**			
Kins-X	0.99	0.99	0.99	0.99	0.38	**0**		
Cars	0.89	0.89	0.94	0.90	0.99	0.98	**0**	
Dunn	0.69	0.7	0.58	0.54	0.99	0.99	0.92	**0**

The values on the top are for the normalised graph edit distances (GED) (non-normalised values can be found in the appendices), followed by the values for the IM distance, then the values for the 1-NMI (NMI values can be found again in the appendices). The smallest values, which represent the graphs that are closest, are emphasised in **bold**. The largest values, which represent the graphs that are the least closest match, are *italics*

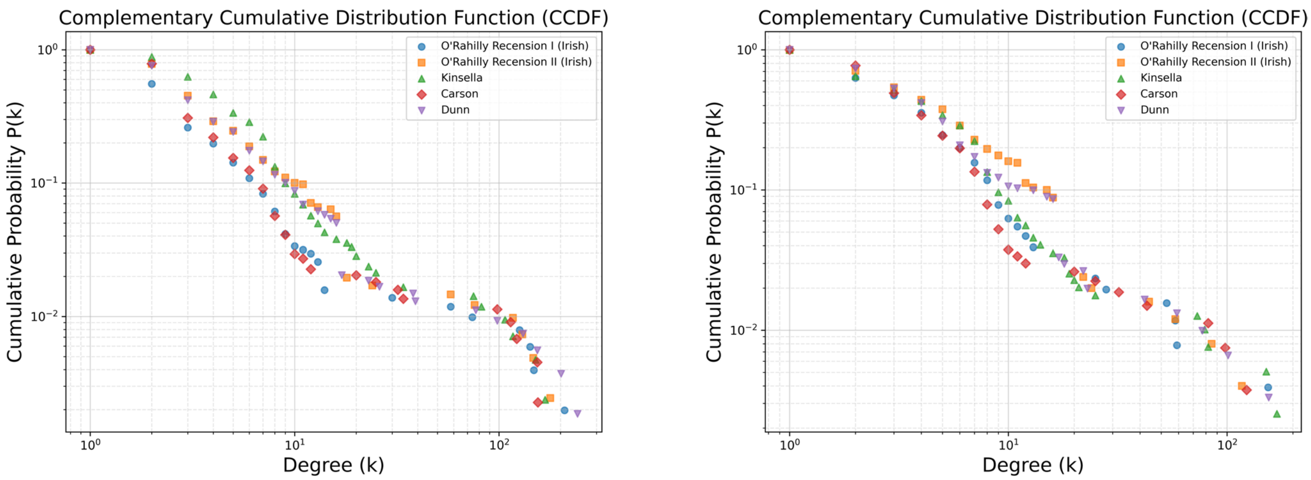

Fig. 1 These figures show the complementary cumulative distribution functions for O'Rahilly, Kinsella, Carson and Dunn's networks. The panel on the left shows the networks which contain the large groups of peripheral characters. The right panel shows the distributions with these groups of peripheral characters removed

Table 3 KS Statistic values when comparing each of the degree distributions to one another

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
ORI-Eng	**0**							
ORI	**0.00**	**0**						
ORII	0.06	0.05	**0**					
OR-comb	0.04	0.04	0.04	**0**				
Kins	*0.13*	0.12	0.1	0.12	**0**			
Kins-X	0.11	0.1	0.09	0.1	0.06	**0**		
Cars	0.09	0.08	0.07	0.08	0.07	0.05	**0**	
Dunn	0.06	0.06	0.03	0.05	0.09	0.07	0.06	**0**

The smallest values, which represent the distributions most similar, are emphasised in **bold**. The largest values, which represent the graphs that are the least closest match, are *italicised*

Table 3 reports the KS statistics for each network compared to the others. A KS statistic value of 0 indicates that the degree distributions are identical, and a value of 1 indicates that they are maximally different. Like the IM distance earlier, all of the values found in the table are below 0.2 as they are all similar distributions. The column with the Kinsella comparisons generally has the largest KS statistic values, which can again be traced back to Kinsella's inclusion of pre-tales. This is confirmed in the adjacent column where the values are all lower when the pre-tales are removed. The Irish and English versions of O'Rahilly's Recension I have the lowest value, meaning their distributions are the most similar.

The KS statistics comparing each of the networks mirror the results found in the graph matching methods. The O'Rahilly networks are all shown to have highly similar degree distributions. This is in contrast to when the O'Rahilly networks are compared to the Kinsella and Carson networks, where the KS values are larger, indicating structural differences.

3.3 Centrality measures

All studied networks have a small number of protagonists who dominate the network properties. This leads to high disassortativity, as there is a small number of characters which are connected to the majority of the other characters which have a low degree. In this section, we compare the top characters by both degree and betweenness centrality.

When comparing the list of characters with the highest degree across the different networks (see in Table 8 in the appendix), the list of characters is largely the same across the translations. Cú Chulainn has the largest degree in almost all networks, apart from Kinsella's network. In the latter, Conchobar is the character with the largest degree, which again can be attributed to Kinsella's inclusion of pre-tales, as they centre primarily around Conchobar and the Ulstermen before *The Táin*. We can also note the presence of Mugain in both O'Rahilly's combined Recension I and II and Dunn's translation as an important character, both in terms of degree value and betweenness centrality. This is explained by her connection to Medb's genealogy tree (and the many familial connections there) found in Recension II, her link to Conchobar, as well as the inclusion of the list of women with whom Mugain meets Cú Chulainn in one of the Boyhood Deeds stories, a list only found in Dunn's and which therefore increases Mugain's number of connections. As one of the only female characters found in these lists of important characters, it is interesting to see what contributes to her inclusion here.

The top ten characters ordered by betweenness are quite consistent in each (see named characters in Table 9 in the appendix). The results are similar compared to the top ten by degree with Cú Chulainn and Conchobar remaining the two highest ranking characters among the different versions. The same characters who had a large degree also have a high magnitude of betweenness. However, betweenness centrality tends to be better at identifying important characters in narratives compared to other centrality measures (Tarasevich et al. 2023; Mac Carron 2014). Conchobar being the highest ranking character in both betweenness and degree for Kinsella's highlights the prevalence of Conchobar in the pre-tales. The top ten characters for O'Rahilly's English and Irish Recension I are almost identical when ranking in terms of betweenness, emphasising how similar these versions are. The top five characters in each is mostly the same, however the next five differ leading to relatively low Jaccard indices as shown in Table 10. This Spearman coefficient shows that the order also isn't highly consistent (also in Table 10).

3.4 Simulations

To account for the different sizes of the generated networks, the larger networks were truncated to match the size of smaller networks by randomly removing nodes until their sizes were equal. This process was repeated 100 times to account for variability and the network properties were averaged across the truncations. It also allows for further conclusions to be made about the similarities of the networks when size differences are not a factor. As the graph edit distance is rather computationally expensive, and the three similarity measures all yield similar results, we just show values for the IM distances in Table 4.

Table 4 shows there are no large changes in the values after the truncations. This indicates that both the IM distance and KS test values are not greatly impacted by network sizes. The largest values can still primarily be found in Kinsella's column, O'Rahilly's English and Irish translations of Recension I are still the most similar. The KS test values have larger changes when using the truncation methods.

4 Conclusions

At the beginning of this paper, we asked whether it is reasonable to use social networks extracted from modern translations and editions, as in Ref. Mac Carron (2014), or whether differences between versions are significant enough to impact the results. By all methods, we find here that a direct translation of the same story, retains the most similar social network structure to the original, but in this case, is not identical.

The epic *Táin Bó Cúailnge* is found in many fragments and in three different recensions spanning over a few centuries, the older two of which are used here. One of the most popular versions of it for a modern audience is Kinsella's

Table 4 IM statistics and KS test values comparing each of the networks after randomly truncating the networks to be equal in size

IM Statistic	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
ORI-Eng	**0**							
ORI	**0.01**	**0**						
ORII	0.07	0.07	**0**					
OR-comb	0.06	0.05	0.09	**0**				
Kins	0.08	0.07	0.12	0.04	**0**			
Kins-X	0.11	0.12	0.08	0.15	*0.17*	**0**		
Cars	0.04	0.04	0.06	0.07	0.09	0.11	**0**	
Dunn	0.04	0.04	0.06	0.04	0.07	0.11	0.02	**0**
KS Test	**ORI-Eng**	**ORI**	**ORII**	**OR-comb**	**Kins**	**Kins-X**	**Cars**	**Dunn**
ORI-Eng	**0**							
ORI	**0.01**	**0**						
ORII	0.06	0.05	**0**					
OR-comb	0.05	0.05	0.04	**0**				
Kins	*0.13*	*0.13*	0.10	0.12	**0**			
Kins-X	0.11	0.10	0.09	0.11	0.07	**0**		
Cars	0.10	0.09	0.07	0.08	0.07	0.06	**0**	
Dunn	0.07	0.06	0.04	0.05	0.10	0.08	0.07	**0**

The smallest values, which represent the graphs that are matched the closest, are emphasised in **bold**. The largest values, which represent the graphs that are the least closest match, are *italicised* version (Kinsella 1970). This has been used in past work (eg. Mac Carron and Kenna (2012)) but that choice was questioned in Ref. Casey (2012) (with some responses in Ref. Kenna et al. (2024)). That version takes elements of both recensions, but here we show that the social network within is different from either of those recensions as well as their combination.

We show that English and Irish versions of O'Rahilly's Recension I are not identical, but they are very similar. Even with only a few nodes and edges different, the graph edit distance requires 267 changes to match these (see Table 6). However, they are still the most similar with very low differences on the graph matching algorithms and structural properties mostly different in the second decimal place or lower.

Part of the reason for this non-exact matching could be due to the difference in structures between the English and Irish languages leading to a difference in the way characters are perceived to be interacting between one another. It can also be attributed to the version of O'Rahilly's translation of Recension I available online including *glosses* within the body of the text. Glosses are textual additions which appear in the margins or between lines in the original manuscripts, and which occasionally introduce new characters or narrative perspectives, but are not part of the main body of the text. These glosses are absent from the online version of O'Rahilly's edition of the text in the original Irish, and the original printed versions of both edition and translation present them only in footnotes. However, for the translation available online, the glosses are incorporated into the main text. Both of these online versions are the most readily available way to access O'Rahilly's work, and are found on the same platform, yet they seem to reflect two different editorial practices when it comes to the inclusion of glosses specifically. The inclusion (or lack thereof) reflects how translation can serve as a place for additions to the source material, as well as the choices made in the process of making a text available to the public. The version combining elements of all three recensions by Dunn, has more interactions than the other networks, but it matches closer to the O'Rahilly editions than the modern ones.

The networks generated from the later and modern renditions of *The Táin*, diverge further from Recension I. The larger distances observed in two modern versions highlight the narrative differences between the old and the new. Although Carson and Kinsella both used O'Rahilly's Recension II as reference, as well as O'Rahilly's Recension I in Carson's case, they result in networks that are structurally less similar to either of O'Rahilly's individual editions. This emphasises how both Carson and Kinsella's versions of the text selectively incorporate pre-tales and sections of Recensions I & II, which leads to the creation of a text that has never been historically validated. This is important to consider if the networks are used to analyse the text and the interactions between the characters within it, but also the time period during which the text was compiled.

From the point of view of the structural network properties, it is worth noting that all versions have similar quantities as shown in Table 1, all being disassortative, with similar shapes of degree distributions and being small world.

This work demonstrates how translations are interpretive and are prone to altering narrative (and hence network) structures, highlighting the importance of linguistic and editorial choices. An implication here is that, as stories evolve and change the social network presented is not entirely consistent. For future research in this area, justification should be provided for which version is used. If performing a quantitative comparison (as in Ref. Mac Carron and Kenna (2012) for example), then perhaps the different versions should be used rather than just one.

Overall, however, these findings suggest that the closer a version stays to an original source, whether in language or narrative structure, the more structurally similar the resulting social network will be. As Kwon et al. Kwon et al. (2009) have shown, even documents that are intended to be universally consistent, such as the Universal Declaration of Human Rights, carry subtle semantic differences across languages. While staying close to the original source leads to more consistent network representations of the tale, it makes the data gathering more difficult. Computational attempts to extract the network, such as using natural language processing or large-language models are mostly trained on English and therefore, in this case, will not be suitable for working on Old or Middle Irish. Using manual extraction, the requirement for those gathering data to be able to read the original language significantly decreases the pool of those able to extract the social network. This is somewhat discouraging, but it does demonstrate a need for crossing disciplinary boundaries in this kind of research and a requirement for experts in both humanities and the sciences to work together.

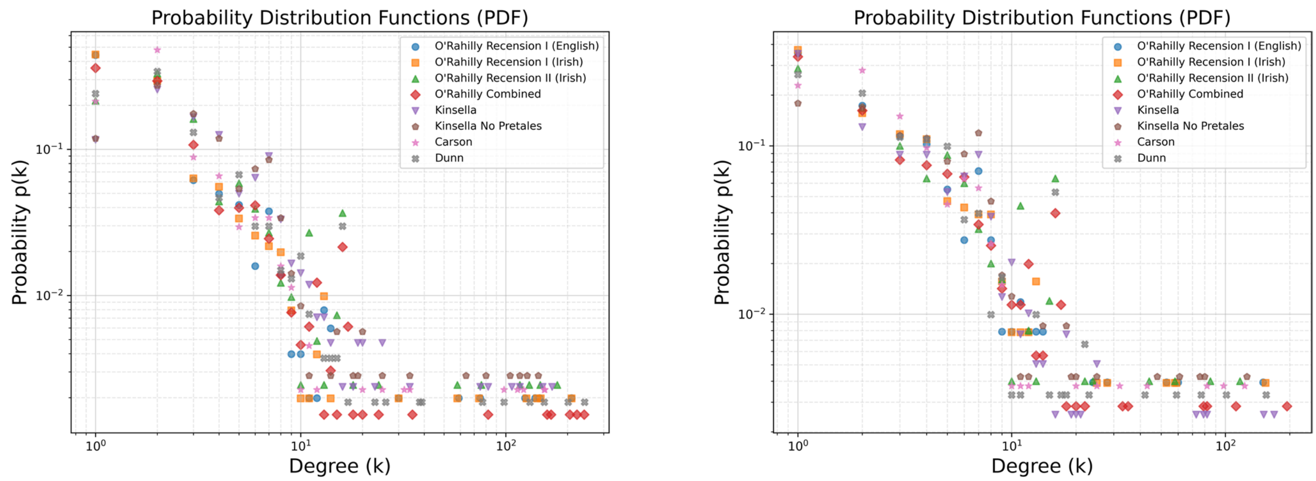

Fig. 2 These figures show the probability mass functions for all networks. The figure on the left contains the large groups of peripheral characters, the right has these groups of peripheral characters removed

Table 5 List of all networks analysed in this paper with abbreviations used to refer to each version henceforth.

Version/edition of the táin	Abbreviation
O'Rahilly Recension I (English)	ORI-Eng
O'Rahilly Recension I (Irish)	ORI
O'Rahilly Recension II (Irish)	ORII
O'Rahilly Recensions I and II Combined	OR-comb
Kinsella	Kins
Kinsella Without Pre-Tales	Kins-X
Carson	Cars
Dunn	Dunn

Table 6 Graph edit distance values when comparing each of the networks, where the most similar networks with the least edits (i.e., the graphs which are most similar) is in **bold**, and the networks with the most required edits (least similar) is in *italics*

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	CARS	Dunn
ORI-Eng	**0**							
ORI	**267**	**0**						
ORII	991	1040	**0**					
OR-comb	1256	1173	833	**0**				
Kins	1571	1560	1704	*1755*	**0**			
Kins-X	1065	1084	1199	1519	1192	**0**		
Cars	969	969	1058	1335	1560	1139	**0**	
Dunn	1171	1115	1159	1372	1738	1367	1113	**0**

Table 7 Network Mutual Information values when comparing each of the networks, where the maximum value is in **bold** (indicating the networks which are most similar), and the minimum value is in *italics* (indicating the most dissimilar networks)

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
ORI-Eng	**1**							
ORI	**0.88**	**1**						
ORII	0.26	0.26	**1**					
OR-comb	0.58	0.63	0.61	**1**				
Kins	0.01	0.01	*0.00*	0.01	**1**			
Kins-X	0.01	0.01	0.01	0.01	0.62	**1**		
Cars	0.11	0.10	0.06	0.10	0.01	0.02	**1**	
Dunn	0.31	0.3	0.42	0.46	0.01	0.01	0.08	**1**

Table 8 Top ten characters in order of highest degree values

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
1	Cuchulainn	Cuchulainn	Cuchulainn	Cuchulainn	Conchobar	Cuchulainn	Cuchulainn	Cuchulainn
2	Medb	Medb	Medb	Medb	Cuchulainn	Medb	Medb	Medb
3	Fergus	Fergus	Fergus	Fergus	Medb	Ailill	Ailill	Fergus
4	Ailill	Ailill	Ailill	Ailill	Ailill	Fergus	Conchobar	Ailill
5	Cethern	Cethern	Conchobar	Conchobar	Fergus	Conchobar	Fergus	Conchobar
6	Conchobar	Conchobar	Cethern	Cethern	Conall	Cethern	Cethern	Ferdiad
7	Finnabair	Docha	Eochu	Laeg	Cethern	Fiacha	Ferdiad	Laeg
8	Ferdiad	Finnabair	Find	Ferdiad	Dubthach	Finnabair	Finnabair	Cethern
9	Docha	Maine Andoe	Findoman	Mugain	Fiacha	Laeg	Maine	Mugain
10	Maine Andoe	Ferdiad	Findeoin	Cormac	Finnabair	Ferdia	Conall	Fiacha

Table 9 Top ten characters in order of greatest betweenness centrality values

	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
1	Cuchulainn	Cuchulainn	Cuchulainn	Cuchulainn	Conchobar	Cuchulainn	Cuchulainn	Cuchulainn
2	Fergus	Fergus	Medb	Medb	Cuchulainn	Medb	Medb	Fergus
3	Ailill	Ailill	Fergus	Fergus	Medb	Ailill	Ailill	Medb
4	Medb	Medb	Conchobar	Ailill	Ailill	Fergus	Conchobar	Conchobar
5	Conchobar	Conchobar	Ailill	Conchobar	Fergus	Conchobar	Fergus	Ailill
6	Cethern	Cethern	Mac Roth	Mac Roth	Eochu	Eochu	Eochu	Ferdiad
7	Bricriu	Bricriu	Cethern	Cethern	Laeg	Ferdia	Cethern	Mac Roth
8	Rochad	Etarcomol	Ferdiad	Ferdiad	Mugain	Laeg	Finn	Donn Cuailnge
9	Etarcomol	Nemain	Dubthach	Mugain	Cathbad	Cethern	Finnoman	Dubthach
10	Nemain	Conall	Docha	Laeg	Finnbennach	Etarcomol	Finnen	Mugain

Table 10 Jaccard and Spearman values for the rankings of characters in order of betweenness, where the most similar rankings are **bold** and the least similar are *italicised*.

Jaccard	ORI-Eng	ORI	ORII	OR-comb	Kins	Kins-X	Cars	Dunn
ORI-Eng	**1**							
ORI	**0.82**	**1**						
ORII	0.43	0.43	**1**					
OR-comb	0.43	0.43	0.67	**1**				
Kins	*0.33*	*0.33*	*0.33*	*0.33*	**1**			
Kins-X	0.43	0.43	0.43	0.43	0.43	**1**		
Cars	0.43	0.54	0.54	0.54	0.33	0.43	**1**	
Dunn	0.43	0.43	0.54	0.82	*0.33*	0.43	0.54	**1**
Spearman	**ORI-Eng**	**ORI**	**ORII**	**OR-comb**	**Kins**	**Kins-X**	**Cars**	Dunn
ORI-Eng	**1**							
ORI	0.67	**1**						
ORII	0.07	0.02	**1**					
OR-comb	0.09	0.18	**0.77**	**1**				
Kins	-0.15	*-0.67*	0.10	-0.08	**1**			
Kins-X	0.60	0.31	0.15	-0.15	0.02	**1**		
Cars	0.39	0.67	-0.01	0.21	-0.19	0.38	**1**	
Dunn	- 0.24	-0.04	0.33	0.67	-0.04	-0.15	0.28	**1**

Appendix A: Additional figures

In this section we present further results in the form of figures. These are the probability mass functions for all versions of the Táin with and without groups.

Appendix B: Additional tables

In this section we present further results and details on the comparison of these networks. Initially however, we show the abbreviations for each of the networks for the reader's convenience in Table 5.

In the results section above we provide the less common normalised graph edit distance. As the graph edit distance is an integer displaying how many changes need to be made to a graph to match another, in Table 6 we show these non-normalised values.

In the same section, we show the network mutual information (NMI) values which are provided in the form 1-NMI. Table 7 displays these values in their natural form.

The top ten characters in each network ranked in terms of their degree are shown in Table 8.

The top ten characters ordered by betweenness are shown in Table 9.

The Jaccard coefficient and Spearman's rank coefficient for the betweenness is shown in Table 10.

Acknowledgements The authors would like to thank the late Ralph Kenna for early discussions on the merits and methods for comparing different versions of the same narrative, and Colm Connaughton for input on graph matching.

Author contributions All authors reviewed the manuscript.

Funding PMC and GC are supported by the Irish Research Council, grant number COALESCE/2024/3349.

Data availability The code and data used for this research can be found at the following GitHub https://github.com/mckermittstephen/Comparing_Tain.

Declarations

Conflict of interest The authors declare no competing interests.

References

Labatut V, Bost X (2019) Extraction and analysis of fictional character networks: a survey. ACM Comput Surv (CSUR) 52(5):1–40

Janickyj M (2023) Textual analysis with natural language processing. PhD thesis. Coventry University Coventry

Elson D, Dames N, McKeown K (2010) Extracting social networks from literary fiction. In: Proceedings of the 48th Annual meeting of the association for computational linguistics. pp.138–147

Mac Carron P, Kenna R (2012) Universal properties of mythological networks. Europhys Lett 99(2):28002

Massey SE (2019) Form and relationship of the social networks of the new testament. Soc Netw Anal Mining 9:1–13

Mac Carron P (2014) A Network theoretic approach to comparative mythology. PhD thesis. Coventry University Coventry

Sanfeliu A, King-Sun F (1983) A distance measure between attributed relational graphs for pattern recognition. IEEE Trans Syst Man Cybern 3:353–362

Ipsen M (2004) Evolutionary reconstruction of networks. In: Function and Regulation of cellular systems , 241–249

Felippe H, Battiston F, Kirkley A (2024) Network mutual information measures for graph similarity. Commun Phys 7(1):335

Conte D et al. (2004) Thirty years of graph matching in pattern recognition. Int J Pattern Recognition Artific Intell 18(03):265–298

Amalvy A et al. (2024) Interconnected kingdoms: comparing 'a song of ice and fire' adaptations across media using complex networks. Soc Netw Anal Mining 14(1):199

Danowski JA (1993) Network analysis of message content. Progress Commun Sci 12:198–221

Carley K (1993) Coding choices for textual analysis: a comparison of content analysis and map analysis. Sociol Methodol 75–126

Doerfel ML, Barnett GA (1999) A semantic network analysis of the international communication association. Human Commun Res 25(4):589–603

Kwon K, Barnett GA, Chen H (2009) Assessing cultural differences in translations: a semantic network analysis of the universal declaration of human rights. J Int Intercult Commun 2(2):107–138

MacCarron P (2024) Quantifying the role of supernatural entities and the effect of missing data in irish sagas. Condensed Matter Phys 27(03):33803

Lei B, Wu Z, Yang H (2025) Uncovering accessible narration underpinning adaption of literary novels based on character network. Int J Modern Phys C 36(11):2550020

Wright JM (2011) A companion to irish literature. Vol.1. John Wiley & Sons

O'Rahilly C (2001a) T á in B ó C ú alnge Recension I. https://research.ucc.ie/celt/document/G301012. Last Checked: 25.06.2025. Original Year

O'Rahilly C (2001b) T á in B ó C ú alnge Recension 1. Ed. by Cecile O'Rahilly. http://research.ucc.ie/celt/document/T301012. Last Checked: 25.06.2025. Original Year

O'Rahilly C (2000) T á in B ó C ú alnge from the Book of Leinster. Ed. by Cecile O'Rahilly. https://celt.ucc.ie/published/G301035/index.html. Last Checked: 25.06.2025. Original Year

Kinsella T (1970) The T á in. Oxford University Press

Carson C (2007) The T á in. Penguin Classics

Dunn J (1914) The Ancient Irish Epic Tale T á in B ó C ú alnge. The Cualnge Cattle-raid,": Now for the First Time Done Entire Into English Out of the Irish of the Book of Leinster and Allied Manuscripts. D. Nutt

Strachan J, O'Keeffe JG (1912) The T á in B ó C ú ailnge from the Yellow Book of Lecan: with variant readings from the Lebor na Huidre. School of Irish learning. Supplement to Ériu (1904-12) vi, 126 p. 25 cm. Dublin and London: Royal Irish Academy

Newman ME (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208701

MacCarron P, Mannion S, Platini T (2023) Correlation Distances in Social Networks. In: Journal of Complex Networks 11.3 (June), cnad016. ISSN: 2051-1329. https://academic.oup.com/comnet/article-pdf/11/3/cnad016/50599857/cnad016.pdf. https://doi.org/10.1093/comnet/cnad016

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world networks'. Nature 393(6684):440–442

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Modern Phys 74(1):47

Newman M (2013) Networks: an introduction. Oxford Scholarship Online, Toronto

Newman ME, Park J (2003) Why social networks are different from other types of networks. Phys Rev E 68(3):036122

Fronczak A, Fronczak P, Hołyst JA (2004) Average path length in random networks. Phys Rev E-Stat Nonlinear Soft Matter Phys 70(5):056110

Donnat C, Holmes S (2018) Tracking network dynamics: a survey using graph distances. Ann Appl Stat 12(2):971–1012

Mannion S, MacCarron P (2023) A robust method for fitting degree distributions of complex networks. J Complex Netw 11.4 (July), cnad023. ISSN: 2051-1329. https://academic.oup.com/comnet/article-pdf/11/4/cnad023/50921127/cnad023.pdf. https://doi.org/10.1093/comnet/cnad023

Clauset A, Shalizi CR, Newman ME (2009) Power-law distributions in empirical data. SIAM Rev 51(4):661–703

Newman ME (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256

Cohen R, Havlin S (2003) Scale-free networks are ultrasmall. Phys Rev Lett 90(5):058701

van der Hofstad R, Komjáthy J (2017) When is a scale-free graph ultra-small? J Stat Phys 169:223–264

Newman ME (2005) Power laws, Pareto distributions and Zipf's law. Contemporary Phys 46(5):323–351

Tarasevich YY, Danilova AV, Romanovskaya OE (2023) Network analysis of verbal communications in the novel the master and margarita by ma bulgakov. Adv Complex Syst 26(01):2350001

Casey D (2012) Textual problems with mathematical assessments of relationship networks in medieval literature (or why the numbers don't add up in Táin Bó Cúailnge). In: academia.edu preprint www.academia.edu/1877262

Kenna R, MacCarron P, Casey D (2024) Comparative network analysis of the Táin Bó Cúailnge. In: 5th International Conference on the ulster cycle of tales. Ed. by David Stifter