Network Theory: Universal Properties of TV Shows
Final report: Math 485

Roopa Krishnaswamy, Jamie Fitzgerald, Manuel Villegas, Riqu Huang, and Riley Neal
Department of Mathematics, University of Arizona, 617 N Santa Rita, Tucson, AZ 85721

Abstract:

Network theory is a growing field that in recent years has been expanded to the analysis of social networks. In this project, five networks, developed from the television programs Cops, The Tudors, Keeping Up With The Kardashians, Deadbeat, and House of Cards, were analyzed using universal properties of networks found by Carron and Kenna in their paper, Universal Properties of Mythological Networks (1). Properties of networks allow the user to determine whether the network presented resembles a real social network or a fictional network. After finding and analyzing the universal properties found in the five television networks, the shows were classified according to their fictitiousness. Overall, it was found that Cops resembled the most realistic social network.

Introduction:

In this project, the first seasons of five television programs, The Tudors, Keeping Up With The Kardashians, House of Cards, Cops and Deadbeat, were analyzed using network theory. Each show's network is comprised of every character-to-character interaction within the season. This interaction is established when one character communicates with another character through various mediums or has a significant impact on another character in the show. By examining these connections, the interconnectedness of the TV show networks will be determined, as well as if the story lines resemble those of realistic social networks or fictional networks. The universal properties presented in the paper by Carron and Kenna (1), who used statistical metrics of network theory to determine the plausibility of networks found in epic literature, will be used to analyze all of the networks for the various TV shows. NetworkX in Python as well as hand calculations will be implemented to determine the values of the various universal properties within each network. Based upon these calculations the plausibility of each network will be determined.

Methods and Tools:

Within Network Theory, methods of statistical analysis are used to classify and compare various networks quantitatively. These connectivity descriptors are calculated typically from nodal degrees, the number of neighbors each element of the network links to. Degree distributions are the probabilistic histograms used when considering the degrees of a network as a whole; these distribution functions yield the probability of a random element having each degree value. Degree distributions functions of real social networks often behave following a power-law dependency on node degrees: $p(k) \sim k^{-\gamma}$, where $p(k)$ is the probability of a node having degree, k and γ is a positive, constant parameter of the network (1).

Scale-free networks are a category of complex networks, frequently occurring in many areas of study and defined by this type of power-law degree distribution, specifically where γ falls between two and three; this distribution is indicative of large, thoroughly connected hub nodes supported by lesser degree node connections (5). Scale-free networks are of interest in that they are characterized as robust networks with small world efficiency. The robustness of a network refers to its resilience, a measure of the extent to which a network remains connected after the removal of random nodes (1). While, small world efficiency is the idea that a relatively short distance must be traveled between any given nodes, resulting in a compact system diameter or width. Graphically, scale-free networks demonstrate giant components formed by influential nodes connecting many other smaller degree nodes, accounting for the robustness of the network (1).

A measure of how influential a node is may be quantified by its Betweenness Centrality, a statistical measure of flow or crucial paths through a network. A geodesic is a path of the shortest length between two elements of a network, so the proportion of all geodesics, σ_l, through a point relative to the total number of geodesics possible, σ, measures how central to the network the point is (1). This calculation is normalized to account for the possibility that all geodesics pass through a certain node and is calculated by the following, where centrality of node-l is denoted, g_l (1):

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

Additionally, network nodes may be characterized by their clustering coefficient, a measure of the density of connections indirectly surrounding them. The clustering coefficient of an individual node is calculated as the percentage of possible connections between immediate neighbors, actually realized within the network (1). Taken as an average over an entire network, the clustering coefficient may be used as a measure of transitivity and yields the probability that any two neighbor-nodes are connected (1). As a percentage, the coefficient will fall between zero and one, those closest to the latter indicating networks that are “cliquey” or strongly grouped. A specific sub-branch of complex and scale-free networks, known as hierarchical networks, are known to exhibit clustering coefficients in this manner (1).

Among network descriptions, the term hierarchical is applied to those in which modular structures form largely isolated groups. Networks containing these node communities are classified as having clustering coefficients that follow a dependency on the inverse of its node degrees (1). Easier demonstrated graphically, hierarchical networks display predominately low-degree nodes organized into communities by few structural high-degree vertices; this results in dense sub-graphs of low connectivity points.

A small world network is a type of network in which most nodes are not connected to each other, but they can be reached by a small number of steps. The way to determine if a network is small world network is to generate another random network that has the same number of nodes and mean degree. Then calculate the average path length and clustering coefficient for both the original network and the random network. If the average path length of the original network is similar to the average path length of the random network, and the clustering coefficient of the original network is much larger than the clustering coefficient of the random network, then it can be determined that the network is small world (1).

Closed triads are often found in scale free networks. A triad is a group of three nodes. A closed triad means every node in the triad has degree of two, so every node in the triad has a connection.

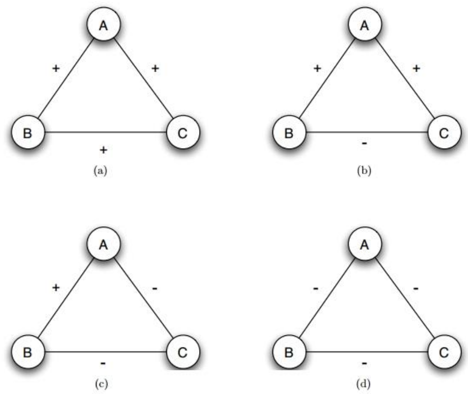


Figure 1. Closed triads.

In a network, a closed triad with an odd number of negative edge is known as structural balance (3). Take a closed triad using friendly and hostile as edges as an example. A positive edge signifies friendly and negative edge signifies hostile. In figure 1(a), nodes A, B, C are mutual friends. The structure of the closed triad is balanced and there are zero negative edges. In figure 1(b), A is friends with B and C, but B and C don't get along with each other. In such a case, A tends to make B and C become friends or A will have to choose sides with either B or C. The structure of the closed triad is not balanced and there is one negative edge. In figure 1(c), A and B are friends with C as a mutual enemy. The structure of the closed triad is balanced and there are two negative edges. Finally, in figure 1(d), A, B, C are mutual enemies. In such a case, two of the three will be motivated to team up against the third, since "the enemy of my enemy is my friend". The structure of the closed triad is not balanced and there are three negative edges. Overall, this propensity to disfavor odd number of hostile links in a closed triad is known as structural balance (4).

Assortativity is a preference for a network's nodes to attach to others that are similar in some way. In many observable networks, it is easy to find correlations between nodes with similar degrees (2). If highly connected nodes tend to connect with other high degree nodes this is called assortativity. If high degree nodes tend to connect with low degree nodes, then such a tendency is called disassortativity. An assortativity coefficient is a measure of assortativity. It is the Pearson correlation coefficient of degree between pairs of linked nodes (2). Positive values of r indicate a correlation between nodes of similar degree, while negative values indicate relationship between nodes different degree. A zero value of r indicates that the network is non-assortative. Fictional networks are usually disassortative and real social networks are always assortative (1). The assortativity coefficient is given by

$$r = \frac{\sum_{jk} jk(e_{jk} - q_j q_k)}{\sigma_q^2}$$

where j and k refer to two nodes and e_{jk} is the joint distribution of remaining degree of those two nodes, q_j and q_k are the distribution of remaining degree for node j and node k. σ_q^2 is the variance of remaining degree for node q (2).

There are two types of attacks in network theory: target attack and random attack. Target attack means removing nodes with the highest betweenness centralities and random attack means removing nodes randomly. There are two properties of the network within attacks: robust and vulnerable. While attacking the network, if the giant component of the network remains big, it is determined that the network is robust. On the other hand, if the network giant component breaks down, the network is vulnerable.

Results:

All five TV shows possess properties of both realistic social and fictitious networks. A summary of the calculations for each network is show in Figure 2.

Network	N	$<k>$	l	$l$rand	l/lrand	$l$max	C	Crand	C/Crand	Gc	Rem Gc	r	gl
Kardashians	41	8.44	2.07	1.91	1.08	4	0.59	0.21	2.81	1.00	0.825	-0.28	0.027
The Tudors	57	5.90	2.55	2.45	1.04	5	0.32	0.088	3.66	1.00	0.756	-0.29	0.028
House of Cards	52	4.30	2.29	3.21	0.71	4	0.49	0.057	8.59	1.00	0.769	-0.43	0.025
Cops	182	3.10	3.92	4.34	0.90	8	0.33	0.022	14.94	0.80	0.753	-0.30	0.011
Deadbeat	102	3.33	2.20	3.64	0.60	4	0.46	0.014	32.75	1.00	0.225	-0.41	0.012

Figure 2. The size, mean degree, mean path length, maximum path length, clustering coefficient, giant component, remaining giant component after Target Attack, assortativity and betweenness centrality for the TV shows, along with the mean path length and clustering coefficient for similarly sized random networks.

Based upon these results, the universal properties of each network can be determined. Of the five TV shows, Cops had the highest number of nodes whereas Keeping Up With The Kardashians (The Kardashians) had the smallest. The Kardashians, however, had the highest mean degree of 8.44, whereas Cops had the smallest, 3.10, indicating that on average each character in The Kardashians had more connections/interactions with other characters. The mean path length for the TV shows were fairly similar in size, excluding Cops, who had the largest mean path length of 3.92. The maximum path length, or the diameter of the network, was the largest for Cops. The mean path length and maximum path length are expected to be larger for Cops given that the size of the network is much greater than the other four networks. The Kardashians had the largest clustering coefficient whereas The Tudors had the smallest. Given that a larger clustering coefficient is indicative of a fictitious network and a smaller clustering coefficient is indicative of a real social network, based upon the numbers it can be presumed that The Kardashians is fictitious and The Tudors is real. The giant component for all TV shows besides Cops was 100%. Cops was the only network with a giant component less than 90% because it had five smaller networks comprised of a few characters that did not interact with the main characters from the show. The assortativity of all five shows were negative, indicating that on the realm of fictitious to real, [-1,1] respectively, all five shows were fictitious and disassortative, with House of Cards being the most fictitious. The values determined for assortativity brought up some controversy because they imply that The Kardashians is the most real social network, whereas it was expected for Cops to be the most real social network. The Cops show was calculated to have the smallest Betweenness Centrality, signifying a real social network, whereas The Tudors was calculated to have the largest.

In order to determine whether or not a network is small world, the average path length, the average path length for a random network, the clustering coefficient, and the clustering coefficient of a random network were examined. To be classified as small world, $l \approx l_{\mathrm{rand}}$ and $C >> C_{rand}$. An issue arose when determining whether or not the five networks were small world because some of the TV shows, such as The Kardashians, had a ratio of l to l_{rand} around 1, however the ratio of C to C_{rand} was very small. Deadbeat had the opposite effect: a small l to l_{rand} ratio and a very large C to C_{rand} ratio. In both cases it was determined that the network was small world. Thus, all five networks were calculated to be small world.

All five networks underwent target attack. The number of characters that were attacked varied for each network. The Kardashians, The Tudors and House of Cards all had the highest degree node removed. Deadbeat had the two highest degree nodes removed. And Cops had the three highest degree nodes removed. Based upon the remaining giant component values seen in Figure 2, where greater than 0.50 is robust and less than 0.50 is vulnerable, it can be determined that of all five networks, Deadbeat was the only show that did not prove to be robust to target attack. This property is indicative of a real social network.

Based upon the graphs of each network it was determined that all five TV shows had hierarchical networks. Hierarchical qualification is based on the inverse relationship between clustering coefficient dependent upon node degree. It can be seen graphically and verified by calculation.

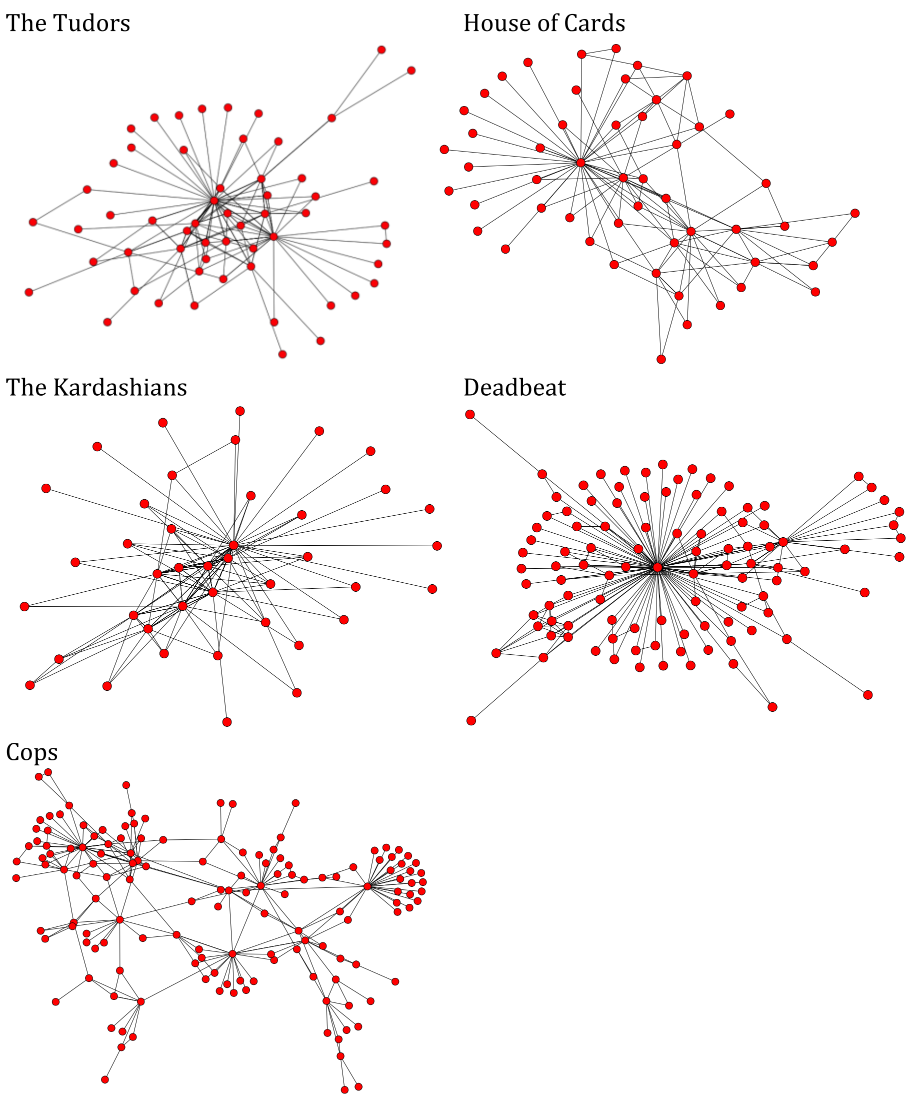
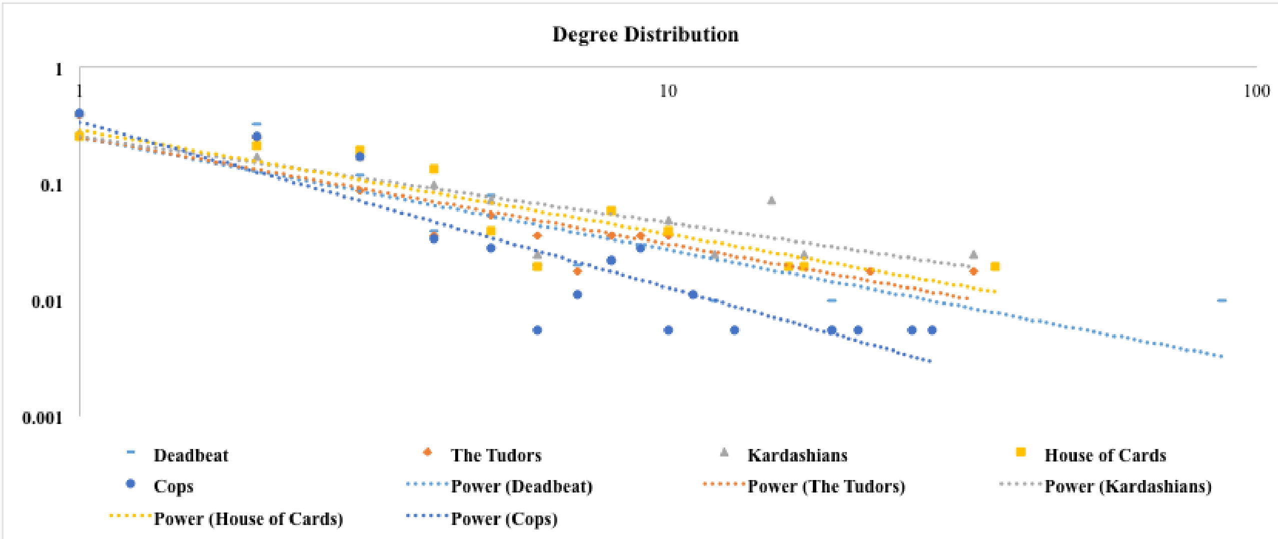


Figure 3. The network corresponding to each TV show.

The degree distributions for each show are plotted in Figure 4.



Network	Equation	R^2
Kardashians	$p(k) = 0.2511k^{-0.739}$	0.73
The Tudors	$p(k) = 0.2491k^{-0.919}$	0.79
House of Cards	$p(k) = 0.2861k^{-0.893}$	0.76
Cops	$p(k) = 0.3367k^{-1.424}$	0.80
Deadbeat	$p(k) = 0.2492k^{-0.971}$	0.74

Figure 4. The degree distribution of each TV show.

Based upon the plot, it was determined that all five networks have a power law distribution. It was also determined that none of the networks were scale free, meaning that none of the networks had a power law distribution with an exponent between -3 and -2. Cops was the closest network to being scale free, with an exponent value of -1.424.

The universal properties of the five TV shows were compared to the real social and fictitious network characteristics found by Carron and Kenna (1). A summary of the results is shown below.

	Social	Fiction	Kardashians	The Tudors	House of Cards	Cops	Deadbeat
Small World	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hierarchy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
p(k)	Power Law	Exponential	Power Law	Power Law	Power Law	Power Law	Power Law
Scale Free	Yes	No	No	No	No	No	No
Gc	<90%	>90%	>90%	>90%	>90%	<90%	>90%
TA	Vulnerable	Robust	Robust	Robust	Robust	Robust	Vulnerable
Assortative	Yes	No	No	No	No	No	No

Figure 5. Summary of properties of TV shows compared to social and fictional networks taken from (1).

Overall, it was determined that all five networks possess qualities of both social and fictitious networks.

Conclusions:

This project conducted analysis on networks formed from character interactions over the first seasons of five different television programs in which the properties of each network were compared against those found in fictional and nonfictional social networks. The resulting networks were all found to display characteristics of being small world networks, forming hierarchical structures with disassortative mixing and exhibiting degree distributions following a power-law dependent upon node degrees. Nonfictional social networks are distinct when compared with fictional networks in that they tend to follow a power law degree distribution with scale free qualities, calculated on a giant component making up less than 90% of the total network and exhibiting vulnerability to targeted attacks. These properties stand in contrast to those of fictional networks that tend to exhibit an exponential degree distribution, calculated on a giant component making up nearly the entire network and displaying robustness under targeted attacks. Additionally fictional networks are typically disassortative, while real social networks show assortative mixing.

To draw conclusions about the degree to which the television programs were fictitious, the properties that differed from one network to the next were compared, and viewed in light of the genre differences represented by each television program. Prior to conducting any testing, it was hypothesized that the show, Cops would display the most properties of a real social network due to its being filmed as a documentary style reality show. Additionally the show Deadbeat was expected to exhibit the most fictitious network properties, as it is known to be a work of purely fiction, not claiming to fall within the realm of reality television. Among the properties used to rank the television shows by similarity to real social networks, the show Cops was confirmed to be the most realistic by it being the closest network to a scale free degree distribution, the only network not made up of a single giant component, and the most vulnerable to targeted attack, with the exception of Deadbeat, a show with a single protagonist with a Betweenness Centrality near 1.

Differences in the nature of the chosen medium between this project and related research, namely television programs versus ancient myths or modern literature can account for the differing in universal network properties as related to the veracity of different stories. The clearest example of this can be seen in this project as the invariance between the qualitative determinations of assortativity; shows with obviously differing realism levels all exhibited disassortative mixing with assortativity *r*-values within close proximity of one another. It was hypothesized that this could be resultant of time constraints on television shows and the inability to show a more encompassing portrayal of the time and interactions of each character.

Additionally the trends in network properties found here, with differences in results or methods from prior research attributed to media formats can be used as reference for further universality comparison within the medium of television shows. Specifically the use of the Betweenness Centrality and remaining giant component percentage following targeted node attacks. The network data showed a correlation Betweenness Centrality and indicators of a fictional network. The show, Cops having the lowest Betweenness Centrality, had the fewest indicators of a fictional network, only showing resemblance to fictional networks under targeted attack and in non-scale-free distribution; and even so, in both measures, Cops was the closest to the qualifications of real social networks when compared against the other programs studied. While the shows with largest Betweenness Centralities (the Tudors and The Kardashians) showed weakest adherence to small world characteristics, the lowest vulnerability to targeted attack, and the most non-assortative mixing between the two of them.

While here these universal properties of networks were examined and measured to speculate on the realistic nature of television programs, potential future application of these methods can be varied in purpose and beneficial to a range of different fields. In contrast to the viewer end of the spectrum, employment of these methods of analysis could be used on the production side to modify and make more realistic television programs, in both the realms of fiction and nonfiction. Outside of the entertainment industry, since veracity of networks show correlations to measures of network interaction, it has been theorized that with a grander range of resources the techniques of analyzing network theory measures could be implemented to the verification of authenticity concerning dealings within legal realms; potentially including the networks of police arrests and crime syndicates, as well as, networks formed by corporate lobbyists and congressmen in collaboration on pieces of legislation.

Acknowledgments:

The authors would like to thank their mentor, George Todd, for all of his help and guidance throughout this project. They would also like to thank Professor Ildar Gabitov for all of his support throughout the semester.

References:

1. Padraig Mac Carron and Ralph Kenna, Universal properties of mythological networks, EPL, 99 (2012) 28002, doi: 10.1209/0295-5075/99/28002
2. M.E.J. Newman, Assortative mixing in networks. *Phys. Rev. Lett.* 89, 208701 (2002)
3. Padraig Mac Carron and Ralph Kenna, Universal properties of mythological networks, EPL, 99 (2012) 28002, doi: 10.1209/0295-5075/99/28002
4. David Easley and Jon Kleinberg, "Positive and negative relationships," in Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010. Print.
5. Anand, Ashish. "Complex Network Theory: An Introductory Tutorial." Department of Computer Science and Engineering. Indian Institute of Technology, Guwahati. 12 Sept. 2013. Lecture.
6. "Season 1." Spacey, Kevin . *House of Cards*. Netflix. 1 Feb. 2013. Television.
7. "Season 1." Seacrest, Ryan. *Keeping Up With The Kardashians*. E!. 14 Oct. 2007. Television.
8. "Season 1." Fellner, Eric. *The Tudors*. Showtime. 1 Apr. 2007. Television.
9. "Season 1." Coady, Chris. *Deadbeat*. Hulu. 9 Apr. 2014. Television.
10. Langley, John, and Malcolm Barbour, prods. "Season 1." *Cops*. 20th Century Fox Television. Los Angeles, California, 11 March 1989. Television.