Social network analysis of character interaction in the Stargate and Star Trek television series

Melody Shi Ai Tan*, Ephrance Abu Ujum† and Kuru Ratnavelu‡

Institute of Mathematical Sciences
Faculty of Science, University of Malaya
50603 Kuala Lumpur, Malaysia

*mel.tan18@gmail.com
†ephrance@gmail.com
‡kuru052001@gmail.com

Received 9 May 2016
Revised 29 July 2016
Accepted 29 July 2016
Published 19 September 2016

This paper undertakes a social network analysis of two science fiction television series, Stargate and Star Trek. Television series convey stories in the form of character interaction, which can be represented as “character networks”. We connect each pair of characters that exchanged spoken dialogue in any given scene demarcated in the television series transcripts. These networks are then used to characterize the overall structure and topology of each series. We find that the character networks of both series have similar structure and topology to that found in previous work on mythological and fictional networks. The character networks exhibit the small-world effects but found no significant support for power-law. Since the progression of an episode depends to a large extent on the interaction between each of its characters, the underlying network structure tells us something about the complexity of that episode’s storyline. We assessed the complexity using techniques from spectral graph theory. We found that the episode networks are structured either as (1) closed networks, (2) those containing bottlenecks that connect otherwise disconnected clusters or (3) a mixture of both.

Keywords: Complex networks; social network analysis; character networks; structure and topology; graph spectra; sociophysics.

PACS Nos.: 02.70.Hm, 89.65.Ef, 89.75.–k, 89.75.Hc.

1. Introduction

Stories trigger our imagination because to some extent we are fascinated by, or are able to identify with, the situations they present. In most stories, a character undergoes changing circumstances through their interactions with other characters. By tracking multiple characters, we can map out character interactions as a kind of

‡Corresponding author.
social network known as a “character network”.\textsuperscript{1,2} Interestingly, recent studies suggest that fiction mimics reality since real world phenomena, such as the small-world effect, manifest in character networks as well. Examples of this include networks in Greek and Roman mythology,\textsuperscript{3} Irish and Anglo-Saxon mythology,\textsuperscript{4} Icelandic mythology,\textsuperscript{5,6} and even the Marvel Comics Universe.\textsuperscript{7}

Television (TV) series, whether fictional or reality-based, are also sources of socially intricate tales and stories, some to the point of becoming the subject of speculative gossip and intense debate amongst its viewers. This is an attempt on the viewer’s part to disentangle the gestalt or mythology surrounding their favorite shows, usually to piece together who has done what with or to whom, and to what end.\textsuperscript{8,9} Every so often, we come across story twists where it is revealed that two seemingly unrelated characters were connected all along; for example in Star Wars V: The Empire Strikes Back when Luke finds out that Darth Vader is actually his father. This recurring theme is referred to as a “trope” which is defined by TVTropes.org as tricks of the trade employed by writers to convey a concept to the audience without explicitly stating the details. The aforementioned trope illustrates how the presence or absence of connections between the right characters has the potential to make a story that much more compelling.

Our contention in this work is that network analysis can be used to discern the complexity of stories in film and TV based on the connectivity of on-screen characters. For a fictional TV series, the stories portrayed in any given episode often revolve around a small-sized main cast, with an intermix of recurring and one-off characters that are linked together according to developments in the plot of that episode, and across the season(s) according to some overarching story arc. As such, the plot must show sufficient continuity and variability from one week to the next in order to maintain or increase their captive audience. Essentially, the variability of the plot manifests itself as a change in the character line-up, or a shift in the dynamic between characters. In this sense, the structure of character networks are to some extent bounded by audience expectations balanced against the writing team’s creative input as well as production constraints. For example, cast and set availability.

In this study, we investigate the connectivity of two TV franchises, namely, Star Trek and Stargate (see Fig. 1). This choice was motivated in part by the size of both franchises (as to supply enough data points for analysis), as well as their similarity in theme and subject matter which can be crudely classified as interstellar space exploration. In particular, we are interested in assessing the complexity of character networks for episodes in both franchises using techniques developed in statistics and spectral graph theory.

2. Reality-Checking Myth and Fiction
An important question when discerning social relations in mythological or fictional works might be framed as: “is the link structure consistent with that found in real social networks?” The answer to this question naturally depends on the
completeness of the network data, since incomplete boundary specification can greatly alter network-level statistics.\textsuperscript{10} This becomes a concern for social networks extracted from a narrative, since characters are mentioned mainly to advance that narrative. Hence, some missing links are to be expected such as those contributed by minor characters, while links contributed by major characters may appear overemphasized. In effect, we are only privy to parts of the story we are told.

If we are to take a character network “as is”, then several comparisons can be made with real social networks (keeping in mind the issue of boundary specification). In the literature, a social network is represented as a graph $G = (V, E)$ with $|V| = N$ nodes (vertices) connected by $|E| = M$ links (edges) and is often found to exhibit the following properties:

- **nonrandom topology.**\textsuperscript{11,12} It is known that large random networks have nodes with number of links (degree), $k$, distributed according to a Poisson probability distribution $P(k) = \lambda^k e^{-\lambda}/k!$, where $\lambda = Np$ and $p$ is the random uniform probability of forming each link.\textsuperscript{13} Most studies seek to find heavy-tailed degree distributions as a sign of complex topology in the sense that the graph in question has structural features somewhere between regular/lattice graphs and random graphs. This intermediacy was originally used to describe “small-world” networks as defined below.

- **“small-world” property.**\textsuperscript{14,15} defined as:
  - being highly clustered like a regular graph or lattice. The local clustering coefficient is defined as $C_i(k) = 2l_i/(k_i(k_i - 1))$ for some node $i$ with $k_i$ adjacent neighbors, where $l_i$ is the number of edges between all of $k_i$ neighbors. For a small-world network, the average local clustering coefficient, $\bar{C} = N^{-1} \sum_i C_i$, is typically orders of magnitude larger than that of a similarly sized random network.
  - having small separation (average path length) between nodes, like that found in random graphs, typically reported to scale as $\langle d \rangle \sim \ln N$ where $\langle d \rangle$ is the average path length.

- **scale-free topology**, although we note at the onset that this is not always the case.\textsuperscript{16} In terms of nonrandom topology, the most prominently discussed in the literature are those in the class of scale-free networks.\textsuperscript{11,17} A scale-free network is characterized by a power-law degree distribution $P(k) \sim k^{-\gamma}$, where $\gamma$ is the scaling exponent, which can be generated by combining network growth and linear preferential attachment.\textsuperscript{18} The Barabási–Albert model corresponds to a rich get richer process,\textsuperscript{19} whereby older nodes have a tendency to connect with newly added nodes with a probability proportional to their degree (i.e. there is an in-built bias for new links to reinforce the connectivity of highly connected nodes). The scaling exponent is typically defined in the range of $2 < \gamma < 3$ for real-world networks.
• **high transitivity** between triples of nodes. If $A$ is connected to $B$ and $B$ to $C$, then $A$ is also likely to have a connection with $C$. This tendency is known as *transitive* or *triadic closure*.20

• **assortative mixing** whereby highly connected nodes tend to connect with other highly connected nodes.21 Specifically, the *assortativity coefficient* is given by the correlation function:

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

where $q_k = (k + 1)p_{k+1}/\sum_j j p_j$ is the *distribution of remaining degree*, $e_{jk} = q_k \delta_{jk}$, and $\sigma_q^2 = \sum k^2 q_k - [\sum k q_k]^2$. The values of $r$ are defined in the range $[-1, 1]$. When $r = 0$ no correlation exists and therefore assortative mixing does not occur, while positive and negative values indicate assortative (similar-degree) and dis-assortative (dissimilar-degree) mixing, respectively. Assortative mixing is a form of homophily whereby similarly connected actors tend to form reciprocated connections,22 or tend to associate to similar social foci.20,23

• **community structure** whereby groups of densely connected nodes are connected by sparse inter-group links.24 This tendency is measured by the *modularity*:

$$Q = \frac{1}{2m} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w),$$

where $A_{vw}$ is an element of the adjacency matrix for the graph in question and is equal to 1 if $v$ is adjacent to $w$ and 0 otherwise. Additionally, $c$ identifies the community label of the node indicated by the subscript while $\delta$ is the Kronecker delta function defined such that $\delta(i, j) = 1$ if $i = j$ and 0 otherwise.2,25 Modularity is defined in the range $[0, 1]$ whereby the farther $Q$ is from zero, the more significant is the community structure (relative to a random graph). Occasionally social networks display such patterns on multiple scales, a trait referred to as *hierarchical structure* and indicated by the presence of a clustering coefficient scaling law $C(k) \sim k^{-\beta}$ with $\beta = 1.26,27$

Previous work on mythological and fictional networks share some overlap in their analysis, and so we compare their main features in Table 1. Interestingly, the mythological and fictional networks listed in Table 1 appear to exhibit small-world behavior — small values of the average path length $\langle d \rangle$, combined with large values of $C/C_r$, where $C$ is the average clustering coefficient of the network and $C_r$ is the average clustering coefficient of a random network of similar size — with a heavy-tailed degree distribution that resembles a power-law scaling. All the networks have a scaling exponent in the expected range of $2 < \gamma < 3$, except for that reported by Alberich and colleagues.7 They pointed out that for $\gamma < 2$ the average properties of the network are dominated by the few highly connected characters in the Marvel Comics universe. Such characters correspond to points at the tail of the power-law distribution which appear as hubs on the network.
Accordingly, hubs play an important role in moderating the fragmentation of a scale-free network under random attack (i.e. hubs provide robustness against the random removal of nodes). However, this also means that scale-free networks are vulnerable to targeted attacks on its hubs. In this sense, we can expect that the removal of a main character on a mythological or fictional network will dramatically affect the remaining structure.

It is also interesting to find that not all of the networks listed in Table 1 display assortative mixing, as expected in real social networks. Gleiser proposed that the presence of disassortative mixing in the Marvel Comics network is a result of the artificiality of the network. In contrast, Mac Carron and Kenna suggested that the degree disassortativity may reflect the adversarial relationship between main characters (i.e. highly connected yet opposing characters are careful not to link too closely with each other).4

Lastly, hierarchical structure appears to be present in the examples covered thus far. This is perhaps to be expected since stories, whether communicated orally, graphically, or by written text, are told as a sequence of events or scenes, each of which usually contains a manageable number of characters (for effective delivery and reception of the story). Accordingly, not all characters get a chance to meet. Low degree characters have a better likelihood of being embedded in complete or near complete nearest neighborhoods (where all of his/her neighbors have also met each other). Conversely, well-connected characters are more likely to contain gaps in their nearest neighborhood at a rate that is approximately inversely proportional to their degree

Table 1. Brief summary of the available works on mythological or fictional networks, that involve topological and structural properties. Small values of the average path length \(\langle d \rangle\), combined with large values of \(C/C_r\), indicate the presence of small-world behavior where \(C\) is the average clustering coefficient of the network and \(C_r\) is the average clustering coefficient of a random network of similar size. The estimated scale-free exponent is denoted by \(\gamma\), whereby the dagger symbol \(\dagger\) indicates a power-law with exponential cut-off. Positive and negative values of the assortativity coefficient \(r\) indicate assortative and disassortative mixing, respectively. The scaling exponent \(\beta\) indicates the level of hierarchical structure based on the scaling of the local clustering coefficient against the degree \(k\), \(C(k) \sim k^{-\beta}\).

<table>
<thead>
<tr>
<th>Type</th>
<th>Source</th>
<th>(\langle d \rangle)</th>
<th>(C/C_r)</th>
<th>(\gamma)</th>
<th>(r)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction7</td>
<td>Marvel Comics</td>
<td>2.6</td>
<td>1.5</td>
<td>0.7(\dagger)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fiction29</td>
<td>Marvel Comics</td>
<td>–</td>
<td>–</td>
<td>2.3</td>
<td>&lt;0</td>
<td>(\approx 1)</td>
</tr>
<tr>
<td>Myth3</td>
<td>Greek+Roman(\text{Unidir})</td>
<td>3.5</td>
<td>120.6</td>
<td>2.6</td>
<td>–</td>
<td>0.6</td>
</tr>
<tr>
<td>Myth4</td>
<td>1. Beowulf(\text{All})</td>
<td>2.4</td>
<td>11.5</td>
<td>2.4</td>
<td>–0.10</td>
<td>(\approx 1)</td>
</tr>
<tr>
<td></td>
<td>(Anglo-Saxon and Greek)</td>
<td>2. Tāin(\text{All})</td>
<td>2.8</td>
<td>41.0</td>
<td>2.2</td>
<td>–0.33</td>
</tr>
<tr>
<td></td>
<td>3. Ísláð(\text{All})</td>
<td>3.5</td>
<td>57.0</td>
<td>2.4</td>
<td>–0.08</td>
<td>(\approx 1)</td>
</tr>
<tr>
<td>Myth5</td>
<td>1. Gísla(\text{Full})</td>
<td>3.4</td>
<td>12.0</td>
<td>2.6</td>
<td>–0.15</td>
<td>–</td>
</tr>
<tr>
<td>(Icelandic)</td>
<td>2. Vatnsvíð(\text{Full})</td>
<td>3.9</td>
<td>16.7</td>
<td>2.7</td>
<td>0.00</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>3. Egils(\text{Full})</td>
<td>4.2</td>
<td>30.0</td>
<td>2.8</td>
<td>–0.07</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>4. Laxdæla(\text{Full})</td>
<td>5.0</td>
<td>25.0</td>
<td>2.9</td>
<td>0.19</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>5. Njáls(\text{Full})</td>
<td>5.1</td>
<td>40.0</td>
<td>2.5</td>
<td>0.01</td>
<td>–</td>
</tr>
<tr>
<td>Combine 1–5(\text{Full})</td>
<td>5.2</td>
<td>100.0</td>
<td>2.8</td>
<td>0.05</td>
<td>&lt;1</td>
<td>–</td>
</tr>
</tbody>
</table>

Accordingly, hubs play an important role in moderating the fragmentation of a scale-free network under random attack (i.e. hubs provide robustness against the random removal of nodes). However, this also means that scale-free networks are vulnerable to targeted attacks on its hubs. In this sense, we can expect that the removal of a main character on a mythological or fictional network will dramatically affect the remaining structure.
degree $k$. Hence, the local clustering coefficient should scale as $C(k) \sim 1/k$ even for mythological and fictional networks.

3. Graph Spectra and Cohesion

Every undirected graph $G = (V, E)$, with $|V| = N$ nodes, is associated to a symmetric adjacency matrix $A$, where $A_{ij} = A_{ji} = 1$ if nodes $i$ and $j$ are connected and zero otherwise. From matrix $A$, we can compute a diagonal degree matrix $D$ with elements $D_{ii} = \sum_j A_{ij}$. Additionally, we can compute the discrete Laplacian matrix $L = D - A$, the eigenvalues of which provide insight into the spectral properties of graph $G$.

Since the Laplacian is non-negative symmetric, we can order the eigenvalues as $0 = \lambda_1 \leq \lambda_2 \leq \cdots \lambda_N$. In particular, the magnitude of the second smallest eigenvalue $\lambda_2$ indicates how well connected $G$ is. This is termed as the algebraic connectivity, where the closer $\lambda_2$ gets to zero the easier it is to bisect the graph into two connected components. The number of zero eigenvalues of $L$ indicates the number of connected components on $G$. If $\lambda_2 = 0$ then $G$ consists of two connected components. If all the eigenvalues are zero then the graph only contains “isolates” (i.e. unconnected nodes). Conversely, the larger is this value the harder it is to introduce a graph cut. Such graphs tend to form closed networks, which have well-connected nodes due to a high overlap in nearest-neighbor connectivity (triadic closure) and extremely short separation (small diameter). In this sense, algebraic connectivity can be used to gauge the cohesion of a network between the two extremes (nearly disconnected or nearly complete).

Another related approach to analyzing network cohesion is from the perspective of a network’s resilience (robustness) to targeted attacks. Mathematically, this instead involves the two largest eigenvalues of the adjacency matrix. Let the eigenvalues of $A$ be arranged as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$. If the spectral gap $\Delta = \alpha_1 - \alpha_2 \approx 0$, then the network contains bottlenecks that, once removed, make the network easier to bisect. On the other hand, if $\Delta$ is sufficiently large (i.e. $\alpha_1 \gg \alpha_2$) then the network is an expander; these are networks that are sparse yet well-connected. Note that the algebraic connectivity can also be used to gauge the robustness or vulnerability of networks to attack.

Estrada and Hatano define a bottleneck as a network formed by two or more clusters of highly interconnected nodes which have very few inter-cluster connections. For our purposes, we adopt this definition to describe the connectivity between two or more subgraphs. Specifically, we describe bottleneck links as bridge links that span densely connected subgraphs. Furthermore, we define bottleneck nodes as node pairs on opposite ends of bottleneck links that span densely connected subgraphs. In this sense, the bisection of a network is easier to attain by the removal of nodes and links that form bottlenecks.

If the first approach to analyzing cohesion uses eigenvalues computed from the graph’s Laplacian matrix $L$, while the second uses its adjacency matrix $A$, a third
approach can be considered using the walk matrix $W$ which is simply the row-normalized version of $A$. This is also commonly referred to as the transition matrix with eigenvalues $1 = \omega_1 \geq \omega_2 \geq \cdots \geq \omega_N$. Intuitively, a graph is well-connected if random walks on that network converge quickly to the stationary distribution. This insight can be used to speed-up the computation of personalized PageRank.36

The second largest eigenvalue $\omega_2$ can determine the mixing rate of random walks on a regular directed graph $G$, with $N$ nodes and walk matrix $W$:

\[
\omega_2(G) \overset{\text{def}}{=} \max_{\pi} \frac{\|\pi W - u\|}{\|\pi - u\|} = \max_{x \perp u} \frac{\|x W\|}{\|x\|},
\]

where $\omega_2(G)$ converges to the uniform distribution $u = (1/N, \ldots, 1/N) \in \mathbb{R}^N$ on $N$ nodes. The first maximum is defined over all probability distributions $\pi \in [0,1]^N$, and the second maximum is defined over all vectors $x \in \mathbb{R}^N$ that are orthogonal to $u$. Random walks on a network with no bottlenecks should mix rapidly. Consequently, a regular directed graph has good spectral expansion for smaller values of $\omega_2$. This insight can be used to gauge spectral expansion on undirected networks; the more regular the topology, the better the spectral expansion, hence the smaller the value of $\omega_2$. Conversely, large values of $\omega_2$ should correspond to networks that deviate from regular topology or contain bottlenecks.

4. Source Data and Network Construction

Character networks were constructed using community-contributed TV episode transcripts for the Star Trek and Stargate franchises, as listed in Fig. 1. In particular, we have relied on transcripts made publicly available at Stargate Wiki38 and GateWorld39 for Stargate, and Chrissie’s Transcripts40 for Star Trek. These transcripts were contributed courtesy of the fans and volunteers who diligently transcribed each episode. We only analyze the SG1, SGA, SGU, TOS, TNG, DS9, VOY and ENT TV episodes. Note that no movies were included because our analysis is only at the episode scale. Movies have larger budgets and cast size that is not comparable to the episodes.

4.1. Structure contributed by each episode

Typically, transcript data for each TV episode adheres to the structure depicted below:

\[
\begin{align*}
\text{SCENE}_1 & \\
: & \\
\text{SCENE}_t & \\
\text{Character}_a &: \text{Spoken dialog} \\
\text{Character}_b &: \text{Spoken dialog} \\
: & \\
\end{align*}
\]
Character\_\textsubscript{\textit{i}}: Spoken dialog

Character\_\textsubscript{\textit{k}}: Spoken dialog

SCENE\textsubscript{\textit{t+1}}

END

This scheme allows us to identify available scenes in a given episode, as well as the characters that appear in it. Accordingly, by parsing through the transcript for some episode \textit{t}, we can construct a weighted undirected graph \( G^{(t)} = (V^{(t)}, E^{(t)}, W^{(t)}) \) such that a weighted link is added between each pair of characters if they fulfill two criteria: (1) they both co-occur in the same scene(s), and (2) they both have spoken lines in the scene(s) they co-occur in. The intended recipient of the dialogue is ignored to simplify the network construction. We refer to such networks here as “episode networks”.

Each scene co-occurrence carries a link weight of 1, therefore if two characters co-occur in \( s \) scenes, then the strength of their interaction in that episode is also \( s \). Essentially, scenes containing \( k \) alternating speaking characters contributes a \( k \)-clique to graph \( G^{(t)} \). Assuming that each episode is associated to a specific plot, then its episode network maps the configuration for that particular storyline. As such, a set of time-ordered transcripts yields a series of character networks from which we can track changes in character connectivity.

![Fig. 1. (Color online) Chronology of the Star Trek and Stargate productions.](image-url)
4.2. Cumulative structure for each series

To better grasp the source data, we also examine statistical regularities in the overall connectivity between characters within each series. This involves the construction of character networks, hereon referred to as “series networks”, in which all TV episode transcripts related to the same series are concatenated into a larger transcript. For example, all 213 episodes of SG1 are grouped together. Similarly, the same reasoning is used for SGA, SGU, TOS, TNG, DS9, VOY and ENT. This allows us to maximize coverage of the transcribed character interactions.

5. Overall Structure and Topology

Table 2 characterizes the properties of the largest connected components for each of the Stargate and Star Trek series network. These giant components, as they are often referred to in the literature, contain the majority of nodes and links. In principle, since a path can be traced between any pair of nodes on the giant component, information can diffuse between the characters represented by them, allowing for domino effects. The transmission of gossip and blame on social networks, for example, can have dramatic implications on an individual’s reputation and trustworthiness.31 Praise can elevate one’s standing, confer more advantages, and open more opportunities. Shame on the other hand, tends to inspire the opposite effect in addition to inciting hostility and aggression. Information diffusion is all the more effective when the giant component has small-world topology. Similar to the studies summarized earlier in Table 1, both Stargate and Star Trek exhibit small-world behavior as signified by their small average path length $\langle d \rangle$ and large average

Table 2. Structural properties of the giant components for the series networks in Stargate and Star Trek. The giant components for the series networks in Stargate and Star Trek comprise ≥ 97% of total nodes and ≥ 99% of total links. The SG series is composed of SG1, SGA and SGU. Likewise, ST is composed of TOS, TNG, DS9, VOY and ENT. The properties measured are the number of nodes $n$, number of edges $m$, average path length $\langle d \rangle$, average clustering coefficient $C$, average clustering coefficient of a random network of similar size $C_{\text{rand}}$, maximum degree $k_{\text{max}}$, power-law scaling exponent $\gamma$, power-law scaling threshold $x_0$, number of points $n_{\text{tail}}$ in the power-law fit, $p$-value for the power-law fit, and assortativity coefficient $r$.

<table>
<thead>
<tr>
<th>Series</th>
<th>$n$</th>
<th>$m$</th>
<th>$\langle d \rangle$</th>
<th>$C$</th>
<th>$C_{\text{rand}}$</th>
<th>$k_{\text{max}}$</th>
<th>$\gamma$</th>
<th>$x_0$</th>
<th>$n_{\text{tail}}$</th>
<th>$p$-value</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG</td>
<td>2026</td>
<td>8840</td>
<td>2.77</td>
<td>0.80</td>
<td>0.004</td>
<td>821</td>
<td>2.6</td>
<td>6</td>
<td>792</td>
<td>0.00</td>
<td>-0.30</td>
</tr>
<tr>
<td>SG1</td>
<td>1433</td>
<td>6088</td>
<td>2.46</td>
<td>0.81</td>
<td>0.006</td>
<td>787</td>
<td>2.8</td>
<td>6</td>
<td>572</td>
<td>0.00</td>
<td>-0.37</td>
</tr>
<tr>
<td>SGA</td>
<td>486</td>
<td>1946</td>
<td>2.38</td>
<td>0.78</td>
<td>0.017</td>
<td>314</td>
<td>2.5</td>
<td>5</td>
<td>221</td>
<td>0.03</td>
<td>-0.39</td>
</tr>
<tr>
<td>SGU</td>
<td>160</td>
<td>853</td>
<td>2.47</td>
<td>0.74</td>
<td>0.067</td>
<td>73</td>
<td>2.4</td>
<td>11</td>
<td>49</td>
<td>0.12</td>
<td>-0.32</td>
</tr>
<tr>
<td>ST</td>
<td>3148</td>
<td>17744</td>
<td>3.01</td>
<td>0.82</td>
<td>0.004</td>
<td>611</td>
<td>2.7</td>
<td>6</td>
<td>1764</td>
<td>0.00</td>
<td>-0.39</td>
</tr>
<tr>
<td>TOS</td>
<td>730</td>
<td>3691</td>
<td>2.28</td>
<td>0.85</td>
<td>0.014</td>
<td>580</td>
<td>2.8</td>
<td>6</td>
<td>403</td>
<td>0.00</td>
<td>-0.38</td>
</tr>
<tr>
<td>TNG</td>
<td>857</td>
<td>4819</td>
<td>2.22</td>
<td>0.85</td>
<td>0.013</td>
<td>608</td>
<td>2.8</td>
<td>6</td>
<td>522</td>
<td>0.00</td>
<td>-0.44</td>
</tr>
<tr>
<td>DS9</td>
<td>697</td>
<td>3841</td>
<td>2.37</td>
<td>0.84</td>
<td>0.016</td>
<td>356</td>
<td>2.1</td>
<td>17</td>
<td>55</td>
<td>0.46</td>
<td>-0.39</td>
</tr>
<tr>
<td>VOY</td>
<td>722</td>
<td>3804</td>
<td>2.37</td>
<td>0.83</td>
<td>0.015</td>
<td>425</td>
<td>2.9</td>
<td>7</td>
<td>327</td>
<td>0.00</td>
<td>-0.46</td>
</tr>
<tr>
<td>ENT</td>
<td>405</td>
<td>1714</td>
<td>2.34</td>
<td>0.80</td>
<td>0.021</td>
<td>285</td>
<td>2.6</td>
<td>5</td>
<td>217</td>
<td>0.00</td>
<td>-0.40</td>
</tr>
</tbody>
</table>
clustering coefficient $C$ (compared to the value attributed to a random network of similar size, $C_{\text{rand}}$). This implies that characters in either franchise are connected by short chains of intermediate acquaintances.\(^{11}\)

As shown in Fig. 2, the series networks for both Stargate and Star Trek exhibit heavy-tailed degree distributions, in which minor characters populate the low-degree range on the left-side of the distribution, while the main characters approach extreme values to the right ($k_{\text{max}}$). This makes good intuitive sense, since minor characters rarely reappear in future episodes unless backed by popular demand. Effectively, this limits the number of links they can accumulate over time. On the other hand, main characters and frequently recurring characters are given more screen time in order to develop their linkages to other characters, both old and new (invariably under extraordinary circumstances), in an effort to cultivate new fans or to deepen the loyalty of existing ones.

Fig. 2. (Color online) The Stargate and Star Trek series network degree distribution. Degree distribution of the series networks in Stargate (blue) and Star Trek (green), using logarithmically spaced bins. The panels are: (a) SG1, (b) SGA, (c) SGU, (d) TOS, (e) TNG, (f) DS9, (g) VOY and (h) ENT. In each panel, the horizontal axis represents degree $k$ while the vertical axis represents probability $P(k)$. Dotted lines represent the associated power-law fit.
Each series network exhibits disassortative degree mixing as indicated by the negative values of $r$. This suggests that well-connected characters tend to link with relatively less-connected characters (and vice versa). Since network growth takes place through the introduction of new characters, who tend to form links with older well-connected characters, this suggests that a power-law fit may be applicable to the networks we have constructed. This is further supported by the presence of hierarchical structure across the board, as indicated by the presence of a scaling law relating the local clustering coefficient and degree, $C(k) \sim k^{-\beta}$ with $\beta \approx 1$ (see Fig. 3). Accordingly, Table 2 includes estimates for the power-law scaling exponent $\gamma$, scaling threshold $x_0$, and the number of points in the fit $n_{tail}$ obtained using the powerlaw Python package readily available online. This package also provides alternative distribution tests based on the methodology outlined by Clauset and colleagues; Klaus and colleagues; and Vuong. In addition, the $p$-value for the power-law fit was estimated using a Python port of Aaron Clauset’s PLPVA code by Joel Ornstein.

Consistent with findings from the Greek and Roman, Anglo-Saxon, and Icelandic mythological networks, the scaling exponent is found in each case to be in the range $2 < \gamma < 3$. However the computed $p$-values for the fit are not found to be significant, except for the cases SGU and DS9 where $p > 0.1$. While this sounds promising, both SGU and DS9 have a large scaling threshold $x_0$ relative to the other series and a $n_{tail}/n < 1/3$. In contrast, 39% to 61% of the nodes make up the power-law scaling range for the other series. Hence even though the $p$-values for DS9 and SGU appear significant, the fit is based on too few data points to yield reliable results. As pointed out, a power-law fit is harder to reject for smaller values of $n_{tail}$.

Consequently, all the Stargate and Star Trek series networks we have constructed cannot be reliably classified as having scale-free topology. Further statistical tests reject alternative distributions in favor of the power-law distribution, either due to insignificant $p$-values (in the case of the truncated power-law distribution) or smaller log-likelihood (log-normal, exponential, and stretched exponential distributions).

The visible gaps in the degree distribution seen in Fig. 2 can be explained by the fact that character connectivity can only be traced from characters with screen time and spoken dialog. As a viewer it is unknown to us to whom each character interacts with behind the scenes, so to speak. In a study of the narrative, the nonlinear order in which a story is told is termed as syuzhet while the reconstructed version according to chronological order is termed as fabula. Since omissions in the story also constitute its syuzhet, the syuzhet sets limits on which characters and links are made known to the audience and which are relegated to “missing data”. This, in turn, affects the goodness-of-fit to known distributions. Accordingly, an elegant fit can only be obtained if the syuzhet features social ties that follow a well-defined probability distribution over a continuum of values for character connectivity (degree $k$).

---

The code is available at http://tuvalu.santafe.edu/~aaronc/powerlaws/.

1750017-11
As previously explained in Sec. 3, a network can be characterized by the extent to which it contains expanders or bottlenecks. We now use this line of reasoning to analyze the episode networks in Stargate and Star Trek. For our purposes, closed networks are indicated by a large maximum coreness $K$, a large spectral gap $\Delta = \alpha_1 - \alpha_2 \gg 0$, and a large spectral expansion $1/\omega_2 \gg 0$. On the other hand, bottlenecks are indicated by a large diameter $D$, high modularity $Q$, and a low

6. Character Cohesion Across Episodes

As previously explained in Sec. 3, a network can be characterized by the extent to which it contains expanders or bottlenecks. We now use this line of reasoning to analyze the episode networks in Stargate and Star Trek. For our purposes, closed networks are indicated by a large maximum coreness $K$, a large spectral gap $\Delta = \alpha_1 - \alpha_2 \gg 0$, and a large spectral expansion $1/\omega_2 \gg 0$. On the other hand, bottlenecks are indicated by a large diameter $D$, high modularity $Q$, and a low

Fig. 3. (Color online) The Stargate (blue) and Star Trek (green) series network local clustering coefficient distribution as a function of degree. The panels are: (a) SG1, (b) SGA, (c) SGU, (d) TOS, (e) TNG, (f) DS9, (g) VOY and (h) ENT. In each panel, the horizontal axis represents degree $k$ while the vertical axis represents the local clustering coefficient $C(k)$. The solid lines (red) represent the associated scaling law relating the local clustering coefficient and degree, $C(k) \sim k^{-\beta}$ with $\beta \approx 1$. 
algebraic connectivity such that $1/(\lambda_2 + 0.1) \gg 0$. For the latter, the addition of 0.1 simply avoids division by zero when $\lambda_2 = 0$.

By combining this information we can represent each episode as a row vector within a feature matrix $X \in \mathbb{R}^{T \times 6}$, where $T$ is the number of episodes considered. Principal component analysis (PCA) can then be used to project $X$ into a set of linearly uncorrelated components. This allows us to: (1) cluster episodes according to their structural similarity in the latent PCA space, and (2) determine how closely, or how far apart, the original features are aligned to each other. The results from this approach are as shown in Fig. 4 for episode networks in Stargate SG1, SGA, and SGU, and in Fig. 5 for Star Trek TNG, DS9, VOY, and ENT. Note that for Star Trek, the PCA plot (Fig. 5) does not include TOS as there were some difficulties that will not be mentioned here. Since the first four PCA components for both Stargate and Star Trek explain 85% of the variance in the feature matrix, $k$-means clustering was used to group episodes according to $k = 4$ clusters. These are color-coded in both Figs. 4 and 5 to aid in identification of cluster membership for each episode.

Fig. 4. (Color online) Stargate episode networks principal component analysis. Principal component analysis using select features of the episode networks in Stargate (SG1, SGA and SGU).
For both plots, we can see that spectral expansion, spectral gap and maximum coreness point to the left of the PC1 axis, while diameter, modularity, and low algebraic connectivity point in the opposite direction (i.e. to the right). We found that episodes to the left of the PC1 axis correspond to closed networks, while episodes located to the right contain bottlenecks. The episode networks also show interpretable variation along the PC2 axis; episodes to the bottom tend to contain a densely connected structure somewhere on its network. Note that episodes closest to the origin do not project exclusively along any particular feature.

Furthermore, three color-coded clusters can be seen around the origin, whereby the central cluster corresponds to episodes that are a mixture of closed network and bottleneck structures. Interestingly, the packing appears to be tighter in the case of Star Trek thus suggesting lower structural variability compared to Stargate. Episodes clustered to the top right, along the “low algebraic connectivity” axis (blue for Stargate and a mix of two colors for Star Trek), are those with $\lambda_2 = 0$. Such episodes indicate the presence of two or more connected components. From a narrative standpoint, fractionalization implies that characters on separate components simply do not meet within that episode. Typically, smaller sized components can be attributed to flashback or flashforward scenes.
For the case of single connected components ($\lambda_2 \neq 0$), bottleneck structures tend to have low algebraic connectivity, high modularity, and occasionally large maximum coreness. In instances where the bisection is particularly good (episodes closest to the “low algebraic connectivity” axis), the resolution of the plot critically hinges on characters spanning the bottleneck links. Meanwhile, characters on the opposite far ends of the bottleneck are not able to interact directly with each other; any interaction is transmitted via intermediaries which is a generic feature of stories involving “hyperlink cinema”. This refers to a style of film wherein characters or action reside in separate scenes but a connection or influence between those disparate stories is slowly revealed to the audience.

This is exemplified in episodes SG1_515 (see Fig. 6(a)) and VOY_306 (see Fig. 6(b)) found along the $Q$-axis in Fig. 4 and the low algebraic connectivity axis in Fig. 5, respectively. Bottlenecks are small sets of nodes whose elimination leads to fragmentations of the network into at least two connected components. The removal of the bottleneck nodes and their corresponding links (see the colored nodes in Fig. 6) fragments the SG1_515 and VOY_306 network into two and four components,
respectively. Typically, storylines that feature bottlenecks tend to focus on the characters that span it. Such characters are in a unique position to control access and timing to information.\textsuperscript{31} Hence, bottleneck structures are ideal for stories involving subterfuge, double-crossing, or moral dilemmas since characters on opposing sides of the bottleneck are not in direct communication with each other.

In the examples given, this is what happens. Episode SG1\textunderscore 515 (Stargate: SG-1 Season 5 Episode 15, titled “Summit”) is about the Goa’uld System Lords meeting on a heavily guarded space station while a truce is in place between them. The Tok’ra plan to take advantage of this rare occasion to kill all of the System Lords using an aerosol-based poison. However, the poison also affects the Tok’ra but not their human allies and therefore they request the help of Dr. Daniel Jackson — who is fluent in Goa’uld — to infiltrate the meeting as a human slave named “Jarren”. Jacob Carter, a member of the Tok’ra acts as Daniel’s handler during his mission helping him whenever possible. While Daniel successfully infiltrates the summit, changing circumstances forces Daniel to delay completing his mission and he must somehow keep his cover while maintaining contact with the Tok’ra. Similarly, in episode VOY\textunderscore 306 (Star Trek: Voyager Season 3 Episode 6, titled “Remember”), Lieutenant B’Elanna Torres receives vivid dreams from an unknown passenger aboard Voyager named “Jora Mirell” about another life, another love and another
planet. Mirell is transmitting those dreams so that the guilt she carries about a past misdeed committed by her people, the Enaran, would not die with her. After Mirell dies, the decision on what to do with those memories rests solely in B’Elanna’s hands; either to pass it on or to keep it to herself. Ultimately, B’Elanna chooses to transfer the memories to another female Enaran, “Jessen”, to keep the memories alive.

For expander structures, we can see that there is a clear distinction between episode networks with a large spectral gap and those with large spectral expansion. In particular, the former is more closely aligned with larger maximum coreness while the latter has a more sparse yet near regular structure. Episode networks structured as expanders typically feature plots that eventually pair up a large fraction of characters. This means that interactions (or knowledge of interactions) between characters tend to diffuse globally and rapidly. In contrast to networks containing bottlenecks, actions and information are globally harder to conceal on expanders since these approximate closed networks. On such networks, bandwidth and the threat of character assassination play a key role in suppressing deviant behavior.

A well-partitioned network makes it possible for individuals embedded in different partitions to act with less “constraint”. Burt defines constraint as situations where a person has contacts with high closure (i.e. are closely connected with one another). Such high closure social circles makes it difficult to conceal actions or information since it may take only one contact to inform all the others, often through gossip. Relating to the concept of bandwidth, Burt likens the network around each person to a broadcast system that is capable of transmitting stories to his/her immediate contacts. The more closure there is within his/her network, the wider (or higher) the bandwidth.

As such, a person’s public image (reputation) depends to some extent on his/her location within a structure of ties. In cases where a person behaves disreputably, others may feel obligated, or presume to take the high moral road and cause indignation to that person. In escalated situations involving conflicting interests, arguments may substitute facts for remarks on the character of that person. Such character assassination becomes even more pronounced with gossip. Closure is a key feature in expander structures; hence, social networks that contain such structures should inhibit the people connected within it from acting in ways that cause reprisals among others. In contrast, bottlenecks may instead have the opposite effect; the segregated structure from a partitioned network creates opportunities for deviant behavior since there is a better likelihood of getting away with questionable actions (relative to networks with high closure).

To illustrate, episodes SG1_404 (see Fig. 7(a)) and DS9_109 (see Fig. 7(b)) are two examples of expander structures found between the spectral expansion and spectral gap axis in Fig. 4 and along the spectral expansion axis in Fig. 5 respectively. Such episode networks require the removal of a sufficiently large number of characters to disconnect it. Since good expanders are highly connected graphs that contain no bottlenecks, it is hard to conceal information in such a closed network.
thus making it difficult for any one character to have power and control over all other characters.

Information in this setting refers to the kind that confers one some kind of strategic advantage over others in his/her network. This could come in the form of a largely unknown weakness or secret that can be exploited, perhaps in a coercive manner. However, such actions are more effective on networks with bottleneck structures whereby bottleneck nodes would have the most advantageous location. On the other hand, the presence of closure on structures such as expanders makes such exploitation less effective since mutual trust and constraint is a common feature on such structures. Closure can be due to triadic or focal closure. However, it is common to find stories that feature high closure networks arising out of distrust, for example, when a group of strangers are stuck with each other but must somehow set aside their differences to achieve some common goal. Even under such settings, each person must in the very least appear to act in accordance to the “best” expectation of all others within that group in order to avoid being singled out and ostracized.

In both episodes SG1_404 and DS9_109, the characters have joined forces to resolve the issue at hand, even when they are strangers to one another. Episode SG1_404 (Stargate: SG-1 Season 4 Episode 4, titled “Crossroads”), is about Shan’auc, one of Teal’c’s old flames and priestess of Chulak. She claims to have discovered how to communicate with her symbiote, Tanith and taught it to despise the Goa’uld. The Tok’ra seem convinced of this, but there is still a reasonable amount of doubt from the others. It is apparently a ruse by Tanith to deceive (by getting a Tok’ra host) and infiltrate the Tok’ra ranks. The Tok’ra is aware of this and allowed this to happen to be better able to control the information Tanith receives. In return, they hope to deceive the Goa’uld, while hopefully gaining further intelligence at some point from Tanith. SG-1 and the Tok’ra are working together to take advantage of
this situation. Meanwhile, episode DS9_109 (Star Trek: Deep Space 9 Season 1 Episode 9, titled “The Passenger”) is about an alien criminal, Rao Vantika attempting to prolong his life, by hiding his consciousness inside the mind of a station crew member, Doctor Julian Bashir. Captain Benjamin Sisko and crew attempt to find and stop Rao Vantika. They are successful by working together with Ty Kajada. In episodes SG1_404 and DS9_109, Tanith and Rao Vantika tried to take advantage of their situation by exploiting what the others did not know about them; portraying false identities of themselves. Ultimately, their actions were unsuccessful due to the high closure of the network.

7. Conclusion

In our investigation of the connectivity of character networks for Stargate and Star Trek, we have made several important observations. First, we found that both Stargate and Star Trek exhibit small-world behavior similar to the mythological and fictional networks summarized in Table 1; signified by their small average path length (i.e. $\langle d \rangle \sim \ln N$) and large average clustering coefficient $C$ (as compared to the value attributed to a random network of similar size, $C_{\text{rand}}$). Thus, two characters chosen at random would most probably be $\langle d \rangle$ hops away from each other, that is characters in either the Stargate or Star Trek franchise are connected by short chains of intermediate acquaintances.41

Additionally, each series network exhibits disassortative degree mixing as indicated by the negative values of $r$ in Table 1. This suggests that older well-connected characters tend to link with newer less-connected characters. The presence of hierarchical structure further supports this finding by the presence of a scaling law relating the local clustering coefficient and degree, $C(k) \sim k^{-\beta}$ with $\beta \approx 1$. It will not come as a surprise since the episode storylines are told as a sequence of scenes or events, the syuzhet, that also introduces the characters and links sequentially.

However, we found no significant support for power-law or truncated power-law scaling over the entire degree range of series networks in either franchise. Due to constraints in screen time, not all links are revealed to the audience which leads to gaps in the degree distribution inferred from all episodes and all seasons in each series. The presence of this incomplete data therefore makes it very difficult to obtain statistically good fits.

Subsequently, we found that most episodes contain a mixture of closed network and bottleneck structures. The former correspond to highly connected graphs while the latter correspond to networks that contain good bisections. A possible explanation for this is that character networks which are between the two types of structure are easier for writers to write, studios to produce, and for audiences to grasp and keep track of. Moreover, such network structures portray the narrative complexity of the plot.47 For example, rejecting the need for plot closure within every episode; feature some episodic plot lines alongside multi-episode arcs and ongoing relationship dramas; oscillate between long-term arc storytelling and stand-alone
episodes; expanding the role of story arcs across episodes and seasons; plot lines are centered upon season-long arcs featuring a particular villain, or “big bad”; and alterations in chronology, e.g. flashbacks, flashforwards, dream or fantasy sequences.

In light of that, we found that “good” spectral expanders as typified by graphs with random walk matrices that yield small values in their second largest eigenvalue are rare, occurring only 5.4% and 5.1% of the time in Stargate and Star Trek, respectively. The outliers are calculated using an interquartile range of 1.5. Comparably, good spectral expanders as typified by graphs with sufficiently large spectral gap are even more rare, occurring only 1.7% and 0.3% of the time in Stargate and Star Trek, respectively. Conflict or tension is usually induced by the introduction or removal of links, so chancing upon the near-optimal ratio relative to a given cast size is difficult. Furthermore, if we consider conflict resolution in stories as a consensus formation process, then expanders have topology that are among the easiest to synchronize globally.

Last but not least, we found that “good” spectral bisections are also rare, occurring only 5.7% and 4.1% of the time in Stargate and Star Trek, respectively. Bisected networks portray clusters of characters with sparse interconnections between clusters. Furthermore, bisected networks can be used to generate a “hyperlink cinema” effect where characters or actions reside in separate scenes. In this case a connection or influence between those disparate stories is only slowly revealed to the audience.

In conclusion, there is still much work that can be done. A limitation to our character networks construction is that it is a static representation where the past, present and future are represented at once. We have ignored the temporality of links when constructing character networks at this time. Indeed, further work can be done to track who met whom on multiple scales, by scene, episode, season and series. Considering that an episode network maps the configuration of the character network for that particular episode plot, that is narratives to networks, it is optimistic that the reverse can be done to generate narratives from networks. Some initial work has already been conducted. However, armed with additional information about the episodes — ratings, reviews, plot, cast, networks, etc. — as training data, we can hope to further improve and build upon the method.

Alternatively, the Stargate and Star Trek character networks provide the opportunity to utilize the large amount of data available to experiment on the robustness of the networks to random or intentional attacks. Expanders, in particular, those with good spectral expansion, are robust against targeted attacks. This is not the case for networks containing bottlenecks, since the network can be fragmented by removing bottleneck nodes; such nodes have high betweenness centrality and low network constraint. We can expect expanders to have some resilience against random attacks since there are a number of redundant pathways to most nodes within that structure. Hence, the removal of any one of these nodes is not likely to fragment the network. Also, networks containing bottlenecks are also not
likely to fragment under random node removal especially when the number of nodes is large.

In future works, it will be interesting to examine how stories are formed through clever use of expander and bottleneck structures. Perhaps in addition to studying how the density or sparseness of linkages constrains the possible stories that can be told, we can also look into the interplay of signed interactions (links) between each pair of nodes. This goes back to studies on social balance theory initiated by Heider to show the balance within the relationship between three things: two persons and an entity that can be an object, situation, event, idea or even another person. This allows us to see how a social group evolves from an imbalance state to a balance state. When a relationship is imbalance, then it will motivate the person(s) involved to right the balance somehow. Heider illustrated that there are four possible balanced states: my friend’s friend is my friend; my friend’s enemy is my enemy; my enemy’s friend is my enemy; my enemy’s enemy is my friend.

By labeling the edges with positive or negative signs, we can convey the relationship between the nodes — whether it represents friendship or enmity. Cartwright and Harary extended Heider’s social balance theory by proving that in a balanced network, either all pairs of nodes are friends, or else the network can be divided into two clusters such that every pair of nodes in each cluster are friends but enemies between different clusters. Thus, we can expect expander structures to have all pairs of nodes to be friends. Networks containing bottlenecks instead would have all nodes on either side of the bottleneck to be friends but enemies for any two nodes from opposite sides of the bottleneck. People of like-minds tend to stick together, that is birds of a feather flock together.

Acknowledgments

The authors thank Michael J. Brunger for useful suggestions and comments. This work was supported by the University of Malaya High Impact Research Grant UM.C/625/1/HIR/MOHE/SC/013 from the Ministry of Higher Education Malaysia. M.S.A.T. acknowledges the financial support under the Malaysian Ministry of Higher Education’s SLAI scheme.

References