

***Avatar: The Last Airbender* Abordagem via Network Science**

Old Bit Waves¹ (alterado para o canal)

¹Programas de Engenharia de Sistemas e Computação
Universidade Federal do Rio de Janeiro (UFRJ)

oldbitwave@gmail.com

Resumo. *Este relatório descreve dois métodos usados para construir e analisar uma rede de personagens da série de televisão Avatar: A Lenda de Aang (como ficou conhecida no Brasil). Os métodos consistem de uma abordagem clássica, implementada em outros trabalhos no passado, como a rede de personagens bem conhecida da obra Les Misérables, e de uma nova, embora ingênua, abordagem cujo foco está nas falas, mais especificamente em vocativos. Além disso, seus respectivos resultados obtidos serão comparados e contrastados a fim de observar as similaridades e diferenças entre as duas abordagens.*

1. Introdução

Diversas obras literárias e midiáticas possuem uma trama referente a um personagem em específico, o dito protagonista. Junto a um protagonista é comum haver a presença de personagens secundários, ou até coprotagonistas, assim como de antagonistas. Em algumas instâncias pode não ser aparente determinar qual personagem é mais central ou importante para a trama, tão pouco determinar o quão mais importante um personagem é em relação a outro.

A série Avatar: A Lenda de Aang (será usada a abreviação ATLA do inglês neste trabalho) consiste de uma trama que possui um conjunto central de personagens junto ao protagonista que a audiência acompanha durante seus episódios. Tal conjunto está em constante interação entre si e com outros personagens da série. Este trabalho procurou analisar se tais propriedades da animação seriam refletidas em uma rede construída apenas a partir dos *transcripts* dos episódios. Tais propriedades seriam refletidas nas características estruturais de uma rede caso o protagonista, coprotagonistas e antagonistas mostrassem ser vértices mais centralizados e de maior importância.

2. Definição do Problema e Objetivo

Tendo em vista a dificuldade de determinar papéis e importâncias em obras literárias, este trabalho tem como objetivo construir, com base nos *transcripts* dos episódios de ATLA, uma rede de personagens. Tendo como base esta rede de personagens, mapea-se as interações entre os personagens da série e analisa se o protagonista se configura como típico de seu papel com base em certas métricas de importância.

Desta forma, procura-se verificar se, por exemplo, o protagonista da série apresenta valores destacados em relação aos demais integrantes da animação. Também deseja-se comparar o quanto os resultados obtidos por meio dos dois métodos propostos neste trabalho se diferenciam. Além disso, almeja-se calcular o coeficiente de Assortatividade [Newman 2018] para cada rede criada a fim de analisar o quanto tais valores refletem a situação da trama da série.

3. Dataset e Modelagem da Rede

Como não há um *dataset* de relações entre personagens e episódios, foi preciso construí-lo com base nos *transcripts* dos episódios. É importante ressaltar que, infelizmente, não há *scripts* originais ou oficiais disponibilizados ao público no momento da realização deste trabalho, logo todos os textos analisados ao longo deste trabalho são propriamente ditos *transcripts*, i.e. redigitação do que foi assistido em um episódio. Devido a estruturação dos *transcripts* [ASN Staff]. E para melhor comparar os dois métodos usados, foi descartado o episódio de índice 215, pois este é o único que não adere ao padrão de escrita em comum a todos os demais episódios (demarcação de Act I, Act II, Act III), contabilizando então 60 episódios analisados no total.

Foram escolhidos dois métodos para mapear os relacionamentos entre os personagens, que serão chamados de Clássico e Leitor neste trabalho. Ambas as redes construídas são não direcionais. O método Clássico se baseia no mapeamento de personagens feito para a obra *Les Misérables* de Victor Hugo realizado em trabalhos anteriores como da Universidade de Buenos Aires [Medus et al. 2005], e do The Koblenz Network Collection [Kunegis 2013] [KONECT]. A priori, o Clássico capturará melhor a aparição de personagens no geral formando cliques, independente de suas falas, enquanto o Leitor terá uma melhor detecção de quanto tempo um personagem ocupa em um episódio. Tendo o enredo em mente, um elemento da narrativa a ser levado em conta para a criação dos vértices é a existência de quatro grupos sociais principais na série. Com isso, também foi observado a relação entre os quatro povos da série. Desta forma, todo personagem da rede possui uma cor respectiva ao seu devido grupo. Abaixo segue uma descrição mais aprofundada de cada método. É importante ressaltar que cada método gerou uma rede distinta, porém ambos os métodos utilizaram do mesmo conjunto de *transcripts*.

3.1. Clássico

O método Clássico consiste em cada personagem ser um vértice, as arestas representarem personagens estarem em um mesmo episódio, e seus pesos o número de episódios compartilhados. Checa-se se o nome de um personagem, ou se uma alcunha correspondente, apareceu no *transcript* de um episódio. Feito isso, todos os personagens presentes neste episódio serão conectados com arestas a todos os outros deste episódio, e o peso de cada aresta representa a quantidade de episódios em comum. Espera-se uma geração de muitos vértices, devido a inclusão de qualquer personagem com nome presente em um episódio.

3.2. Leitor

O método do Leitor consiste também em cada personagem ser um vértice, porém as arestas indicam um personagem chamar o outro pelo seu nome (ou alcunha), e seus pesos o número total de cenas em que isso ocorreu. Se *A* chama *B* ou se *B* chama *A*, codifica-se a mesma aresta. Como os *transcripts* possuem falas e elementos de narração demarcados de maneiras distintas, utilizou-se de tais elementos para demarcar o início e término de uma cena, assim como o falante de cada fala. Para este trabalho define-se cena por toda e qualquer fala entre um elemento de narração que possui um marcador de cena e o seguinte elemento com um marcador similar. Um marcador de cena é um elemento como "*Cuts*", "*Scene changes*", "*Fade to*", "*Commercial Break*", etc. Checa-se se o falante utilizou um nome ou alcunha de outro personagem como vocativo. Foi utilizada uma aproximação para determinar se é um vocativo de modo que se um nome possui um ponto

logo após sua aparição, ele foi considerado um vocativo. E.g. "Katara, [...]", "Katara!", "[...] Katara?", "[...] Katara.". Desse modo, foi checado se ao menos uma vez na cena um personagem invoca outro pelo nome.

4. Métricas Analisadas

Definiu-se a "importância" ou "influência" de um vértice de uma rede pelas métricas de Centralidade de Grau, *Closeness*, *Betweenness*, pesos de arestas, pois estas representam uma medição da interação entre os personagens da série. O Grau de cada personagem indica com quantos personagens distintos um personagem está relacionado. *Closeness* indica o quão central um personagem é em relação aos demais. *Betweenness* denota o quanto um personagem conecta personagens não conhecidos entre si. Os pesos das arestas indicam a quantidade de vezes que dois personagens se relacionam na rede, e portanto é um indicador de prevalência na série. Por fim, foi calculada a medição de Assortatividade de cada rede, que indica relacionamentos entre indivíduos baseados em suas classes (povos).

A priori, independente do método utilizado, é de se esperar que o protagonista possua maior Grau, *Closeness*, e *Betweenness*, mas também que coprotagonistas e antagonistas possuam estas métricas elevadas em relação aos demais personagens. Como o protagonista e coprotagonistas acompanhantes estão presentes na maioria dos episódios, também espera-se que as arestas entre eles estejam entre as arestas de maiores pesos.

5. Resultados

Nesta sessão serão discutidos os resultados obtidos e como eles refletem a dinâmica da série. É de interesse deste trabalho a observação dos valores atrelados aos seguintes personagens: protagonista *Aang*; coprotagonistas *Sokka*, *Katara*, *Appa*, *Momo*, *Zuko*, *Iroh*, *Toph*; e antagonistas *Ozai* e *Azula*. Os valores também foram normalizados, com exceção dos pesos das arestas.

5.1. Propriedades Gerais

A rede obtida pelo Clássico é muito mais densa em comparação à obtida pelo Leitor. Isso se torna evidente ao observar como a centralidade de grau média do Clássico é quase quatro vezes maior que a do Leitor. Também pode-se observar como a centralidade de grau mediana é mais elevada no Clássico, configurando uma rede com mais conexões entre os nós.

Metric	Mode	Min	Max	Mean	Median	Std. Dev.
<i>Centrality</i>	Clássico	0,044	0,977	0,246	0,200	0,214
	Leitor	0,012	0,583	0,055	0,037	0,081
<i>Closeness</i>	Clássico	0,506	0,978	0,581	0,556	0,101
	Leitor	0,253	0,678	0,403	0,414	0,071
<i>Betweenness</i>	Clássico	0,000	0,084	0,006	0,000	0,017
	Leitor	0,000	0,548	0,019	0,000	0,068
<i>Edge Weights</i>	Clássico	1	58	2,883	1	5,543
	Leitor	1	149	5,521	2	14,528

Tabela 1. Resumo das Métricas

Além disso, pode-se notar pelos valores de *Betweenness* mais elevados na rede do Leitor como esta rede é mais centralizada. Isto reflete bem como os dois métodos se divergem, já que o protagonista está presente em 58 dos 60 episódios analisados, o Clássico acaba por conectar a esmagadora maioria dos personagens diretamente a ele, enquanto que o Leitor produz uma configuração mais natural entre os personagens, apesar de não mapear por completo todas as interações possíveis em um diálogo.

Ao comparar os dados relacionados aos pesos das arestas da Tabela 1, pode ser observado como a variação é muito maior no Leitor. Isto mostra como quaisquer que sejam os personagens com arestas mais pesadas, tomam maior partido na trama da história, e, por conseguinte, maior tempo e cenas nos episódios.

5.2. Importância dos Personagens

Para melhor destacar esta diferença, foi montado um mapa gráfico da rede de cada método. Os pesos das arestas e os nomes dos personagens não estão representados graficamente na tentativa de facilitar a visualização da rede. Em ambas as redes os vértices estão coloridos de acordo com os quatro povos da série:

- Azul: Tribos da Água e Pântanos
- Branco: Nômades do Ar
- Vermelho: Nação do Fogo e Colônias
- Verde: Reino da Terra
- Amarelo: (Caso Especial) personagens que não se encaixam em nenhuma das categorias anteriores (espíritos, animais, personagens de origem ambígua) e.g. Appa e Momo.

Pode-se notar como a Figura 1a possui muito mais arestas que a Figura 1b. Em ambos os casos, o protagonista situa-se no centro da rede e cercado pelos coprotagonistas e antagonistas. Também pode-se notar como se caracteriza a Assortatividade entre os personagens ao visualizar as redes, a rede da Figura 1b aparenta juntar personagens de cores, e por conseguinte povos iguais, o que já se torna difícil de observar na rede da Figura 1a.

Os coeficientes de Assortatividade para cada rede são 0,048 para o Clássico e 0,258 para o Leitor, o que demonstra como a primeira rede tende a agrupar personagens não obstante suas origens, o que é esperado já que durante os diversos episódios da série há a presença de inúmeros personagens nos *transcripts*, enquanto a segunda possui uma certa preferência por agrupar personagens de mesma origem, mostrando como personagens tendem a interagir mais com personagens de afiliação igual às suas.

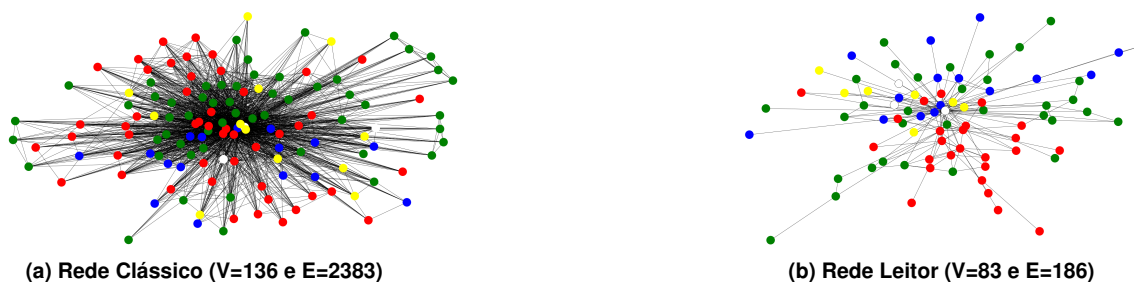
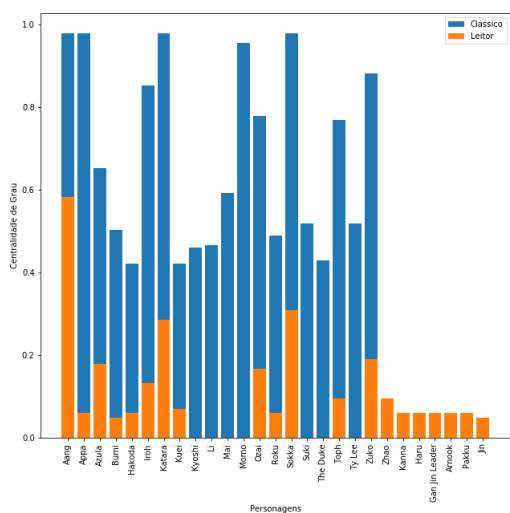
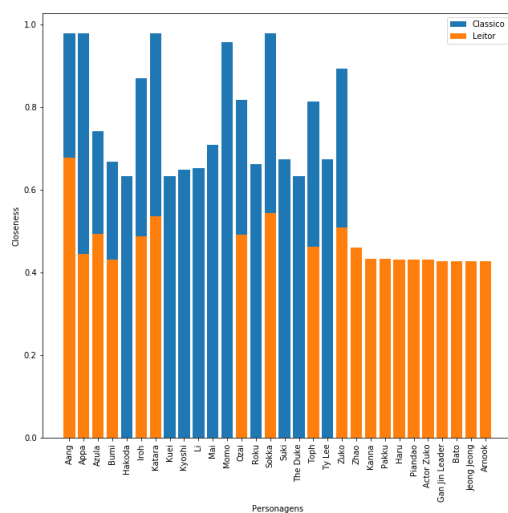


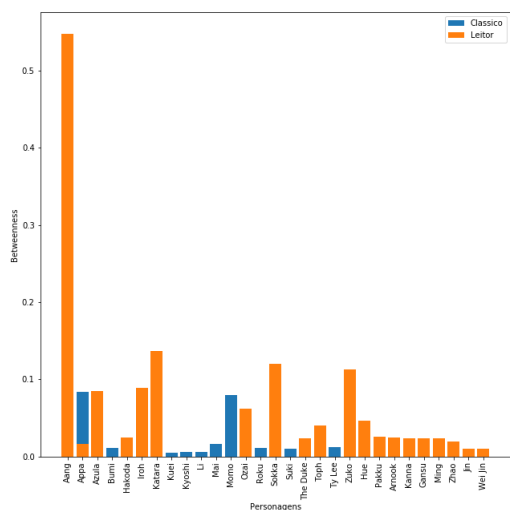
Figura 1. Redes geradas pelos métodos



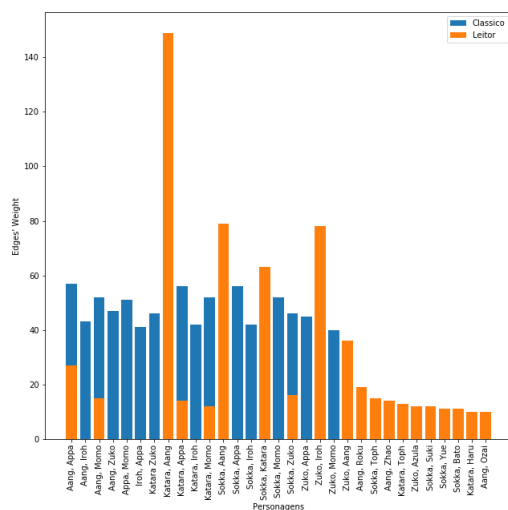
(a) Centralidade de Grau



(b) Closeness



(c) Betweenness



(d) Pesos das Arestas

Figura 2. Métricas - Top 20 personagens mais frequentes

Pelos gráficos da Figura 2a, Figura 2b, Figura 2c, e Figura 2d demonstra-se como todos os personagens centrais para a série estão no Top 20 em ambos os métodos. Apesar disso, o Top 20 começa a se diferenciar mais, devido às características sociais de cada personagem, i.e. quem eles conhece, como e com quem se comunicam. Pode-se ver claramente como o protagonista possui a maior importância em todas as métricas. Os coprotagonistas e antagonistas ainda possuem importância elevada em relação aos demais personagens da rede, entretanto esta é bem menor que a do protagonista.

Observa-se também pelos gráficos acima como a *Betweenness* aumenta no Leitor em relação ao Clássico, o que é de se esperar visto que a rede do Leitor é mais centralizada e possui menos caminhos entre os vértices. De forma análoga, nota-se como algumas interações entre certos personagens passam a possuir pesos mais elevados. Tais arestas possuem personagens que de fato compartilham muito tempo de cena durante os episódios, logo tais diálogos levaram ao grande aumento de determinadas arestas, e quando há pouco diálogo da parte de um personagem, isto é refletido em um peso menor.

Pode-se ver claramente pela Figura 3 como o modelo Clássico gera uma rede com distribuição de graus muito maior que o modelo do Leitor.

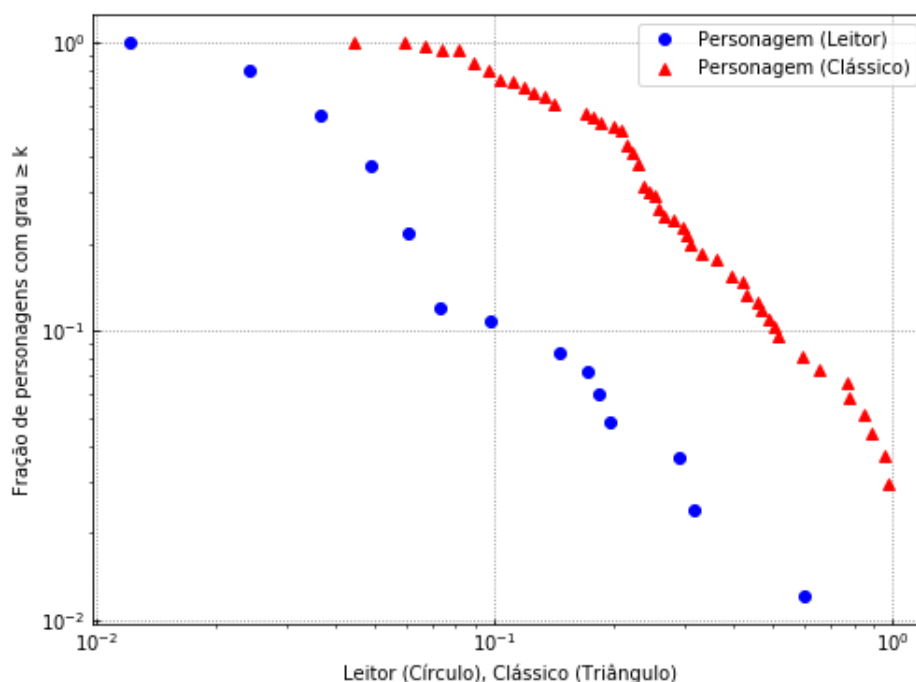


Figura 3. CCDF - Grau dos Vértices

Na Figura 4a e Figura 4b, os pontos representam personagens, e os pontos vermelhos representam os personagens previamente ditos principais para a série. Pela Figura 4a, fica evidente a região em que um ponto é identificado como principal para a obra, já na Figura 4b, há dois pontos em específico que não se destacam, justamente por serem personagens que não possuem falas, e por tanto são penalizados pelo modelo do Leitor; a exceção destes, continua muito evidente um limitante entre personagens principais e não principais.

Além disso, na Figura 4b pode-se notar como alguns pontos pretos estão mais fortes (maiores) do que outros, devido a sobreposição de diversos personagens que mostram possuir características semelhantes nesta rede.

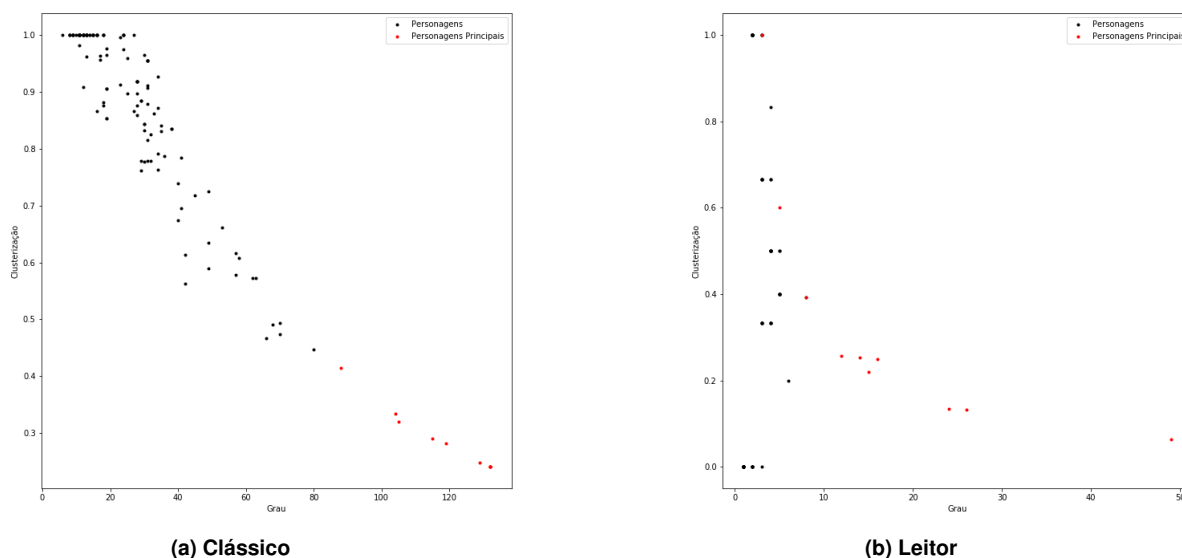


Figura 4. Clusterização x Grau de Personagens

6. Conclusão

Pode-se observar como o protagonista de ATLA se destacou como mais importante independente do método utilizado, Clássico ou Leitor (proposto pelo autor deste trabalho). De fato, os dez personagens mais importantes em todas as categorias de ambos os métodos são em sua maioria os mesmos (protagonista, coprotagonistas, antagonistas), com apenas a ordem de *ranking* ligeiramente diferente. Contudo ao levar em consideração os vinte personagens mais importantes, a elencagem de cada método se torna mais destoante. No fim, a execução do trabalho permitiu concluir como as características estruturais da rede conseguem trazer significados para caracterizar os personagens principais das redes construídas a partir dos *transcripts* da série.

Seria interessante analisar se um aperfeiçoamento do Leitor para considerar uso de pronomes pessoais geraria dados semelhantes. Da mesma forma, seria interessante realizar um estudo parecido para outras obras literárias a fim de determinar as características observadas ao construir a rede de ATLA são reproduzíveis em obras diferentes, ou se por acaso estas características são intrínsecas à esta série.

Referências

- ASN Staff . Transcripts. <http://atla.avatarspirit.net/transcripts.php> . Online; accessed 20-december-2019.
- KONECT . Les misérables. http://konect.uni-koblenz.de/networks/moreno_lesmis . Online; accessed 20-december-2019.
- Kunegis, J. (2013). Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350. ACM.
- Medus, A., Acuña, G., and Dorso, C. O. (2005). Detection of community structures in networks via global optimization. *Physica A: Statistical Mechanics and its Applications*, 358(2-4):593–604.
- Newman, M. (2018). *Networks*. Oxford university press.