

UNIVERSIDADE ESTADUAL DE PONTA GROSSA

Setor de Ciências Exatas e Naturais

Programa de Pós-Graduação em Ciências - Física

PEDRO JEFERSON MIRANDA

EMERGÊNCIA E FLUXO DE INFORMAÇÃO EM REDES COMPLEXAS

Ponta Grossa, PR

2014

UNIVERSIDADE ESTADUAL DE PONTA GROSSA

Setor de Ciências Exatas e Naturais

Programa de Pós-Graduação em Ciências - Física

PEDRO JEFERSON MIRANDA

EMERGÊNCIA E FLUXO DE INFORMAÇÃO EM REDES COMPLEXAS

Dissertação apresentada ao Programa de Pós-Graduação em Ciências, área de concentração Física, da Universidade Estadual de Ponta Grossa, como parte dos requisitos para obtenção do grau de Mestre em Ciências.

Orientador: Giuliano Gadioli La Guardia

Ponta Grossa, PR

2014

**Ficha Catalográfica**  
**Elaborada pelo Setor de Tratamento da Informação BICEN/UEPG**

Miranda, Pedro Jeferson  
M672      Emergência e fluxo de informação em  
            redes complexas/ Pedro Jeferson Miranda.  
            Ponta Grossa, 2014.  
            111f.

            Dissertação (Mestrado em Ciências -  
            Área de Concentração: Física),  
            Universidade Estadual de Ponta Grossa.  
            Orientador: Prof. Dr. Giuliano Gadioli  
            La Guardia.

            1.Redes complexas. 2.Caminhada  
            aleatória. 3.Mitologia comparada.  
            4.Tolerância oral. I.La Guardia, Giuliano  
            Gadioli. II. Universidade Estadual de  
            Ponta Grossa. Mestrado em Ciências. III.  
            T.

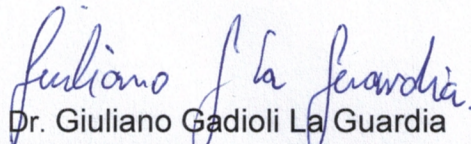
CDD: 519.2

**TERMO DE APROVAÇÃO**

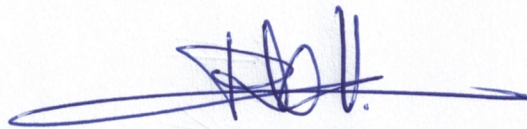
**PEDRO JEFERSON MIRANDA**

**“EMERGÊNCIA E FLUXO DE INFORMAÇÃO EM REDES COMPLEXAS”**

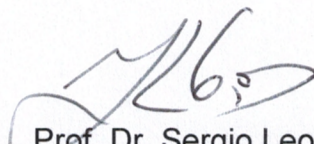
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Ciências - Física da Universidade Estadual de Ponta Grossa, pela seguinte banca examinadora.



Dr. Giuliano Gadioli La Guardia  
Departamento de Matemática e Estatística - UEPG/PR..



Prof. Dr. Ricardo Luiz Viana  
Departamento de Física – UFPR/PR



Prof. Dr. Sergio Leonardo Gómez  
Departamento de Física – UEPG/PR

**Ponta Grossa, 03 de setembro de 2014.**

# Agradecimentos

- A minha mãe, Isabel, que sempre tem me apoiado.
- Ao professor Sandro, por confiar em meu potencial e ter paciência comigo.
- Ao professor André Vitor, por te me apresentado ao grupo de pesquisa DNLSC.
- A todos do grupo de pesquisa que me ajudaram em algum momento no decorrer do curso.
- A CAPES e Fundação Araucária pelo auxílio financeiro.
- A todas as pessoas que estiveram ao meu lado durante o tempo deste estudo.

*"É difícil dar uma ideia global do vasto campo da matemática moderna. A palavra "campo" não é mais adequada: tenho em mente um espaço fervilhando de belos detalhes, não a vastidão uniforme de uma planície nua, mas uma região de um belo país, inicialmente vista a distância, mas merecedora de ser examinada de um extremo a outro e estudada em seus mínimos detalhes: seus vales, rios, montanhas, florestas e flores".*

Arthur Cayley

## RESUMO

A emergência é fenômeno que dá unidade qualitativa a qualquer substância, constituindo o reflexo no ato ontológico da percepção. É a chave conceitual que justifica o uso do modelo em redes complexas para descrever sistemas, que também são complexos naturalmente. Dada essa chave conceitual, buscou-se utilizá-la na geração de novas análises. Para tanto é empregado a teoria de grafos e a caminhada aleatória em dois estudo de caso. Um deles constitui a análise de uma rede mitológica referente à Odisseia de Homero. Foi verificado que a rede mitológica apresenta padrões de redes sociais reais quando excetuados da rede as personagens mitológicas. Em segundo lugar, foi realizado um estudo da tolerância oral como um fenômeno de rede complexa, foi utilizada a caminhada aleatória como modelo estocástico de difusão de estímulos numa rede complexa. Com isso, foi possível conhecer a importância relativa de cada componente imunológica. Por fim, fica evidenciado que o conceito chave de emergência permite a concepção de novas formas de análise, fundamentalmente no uso de redes complexas como modelos que albergam a complexidade inerente na concepção de sistemas reais.

**Palavras-chave:** redes complexas, caminhada aleatória, mitologia comparada, tolerância oral.

## ABSTRACT

The emergence is a phenomenon that gives sense to the qualitative unity of any substance, consisting the reflex in the ontological act of perception. It is the conceptual key that justifies the use of complex network models to describe systems, which also are complex in nature. Given this key concept, it was desired to apply it on real objects in order to create new analysis methodologies. For this, graph's theory and random walk's theory were used as fundamentals for two study cases. One of them consists on an analysis of the mythological social network of Odyssey of Homer. It was found that this network displays structural characteristic of real social network mixed with fictional aspects associated to mythological characters. Another study was the oral tolerance phenomenon modeled as a complex network associated with stochastic dynamics. We applied the random walk as a way to understand the relative importance of each immunological component. Finally, it becomes evidenced that the key concept of emergence allows new forms of analysis using complex network theory as a model which comprises the complexity inherent on the conception of real systems.

**Key-words:** complex networks, random walk, comparative mythology, oral tolerance.



# Lista de figuras

|  |      |
|--|------|
| 2.1 Exemplo de um grafo.....   | p.21 |
| 2.2 Exemplo de um dígrafo.....   | p.32 |
| 2.3 Esquema das Pontes de Königsberg e seu grafo associado.....  | p.36 |
| 2.4 Exemplo de processo de criação de grafos por meio do modelo de Erdos-Renyi.....  | p.39 |
| 2.5 As probabilidades limiares em que diversos subgrafos emergem num processo de evolução de grafo randômico.....  | p.43 |
| 2.6 O modelo de Watts-Strogatz como um aumento no número atalhos rede proporcional à probabilidade de reconecção $p$ .....   | p.44 |
| 3.1 Exemplo de uma cadeira de Markov.....  | p.57 |
| 4.1 Rede mitológica da Odisseia.....   | p.65 |
| 4.2 Gráfico que demonstra a organização hierárquica da rede mitológica da Odisseia. São plotados os coeficientes de aglomeração médios por grau contra os graus..... | p.67 |
| 4.3 Gráfico que expõe a distribuição de graus como uma lei de potência.....  | p.69 |
| 4.4 Gráficos log-log de distribuições de graus (A) para Odisseia, (B) e (C) para a rede de postagens do Facebook.....  | p.70 |
| 4.5 Comunidade referente aos deuses do olimpo e seus descendentes, assim como as entidades mitológicas contidas na Odisseia.....                                     | p.73 |
| 4.6 Comunidade referente aos heróis que participaram na Guerra de Troia e seus agentes diretos. Também constitui a família de Menelau.....                           | p.73 |
| 4.7 Histograma que apresenta a sequencia dos graus máximos de cada comunidade em uma ordem decrescente.....  | p.74 |
| 4.8 Ilustração de uma comunidade em suas duas acepções possíveis.....  | p.75 |
| 5.1 Rede de relações imunológicas associada ao fenômeno da tolerância oral.....  | p.83 |
| 5.2 Representação gráfica dos erros relativos médios de cada nocaute calculados analiticamente (em azul) e algoritmicamente (em vermelho).....                       | p.94 |

## Lista de tabelas

|  |       |
|--|-------|
| 2.1 Matriz de adjacência do grafo da figura 3.1.....   | p.26  |
| 2.2 Matriz de incidência do grafo da figura 3.1.....   | p.27  |
| 2.3 Matriz de adjacência digrafo da figura 3.2.....  | p.32  |
| 4.1 Sumário das propriedades estruturais de diversas redes mitológicas.....  | p.66  |
| 4.2 Ataques direcionados e aleatórios como respostas na componente gigante $G_c$ , em termos do valor absoluto e relativo.....                       | p.68  |
| 4.3 Remoção acumulada dos personagens mitológicos da Odisseia e seu impacto na assortatividade de grau e tamanho da componente gigante.....          | p. 71 |
| 4.4 Sumário das propriedades estruturais de cada comunidade mais importante da rede da Odisseia.....   | p.76  |
| 5.1 Lista das componentes imunológicas que são os vértices da rede ilustrada na figura 5.1.....  | p.81  |
| 5.2 Valores dos erros relativos para os métodos analítico e algorítmico para cada nocaute da rede de interações imunológicas da tolerância oral..... | p.91  |

# Lista de abreviações e siglas

AIR – Resposta imune adaptativa;

APC – Células apresentadoras de antígeno;

MHC – Grande complexo de histocompatibilidade;

TCR – Receptores de células T;

TREG – Células T regulatórias;

nTREG – Células T regulatórias naturais;

iTREG – Células T regulatórias induzidas;

GALT – Tecido linfoide associado ao intestino;

LP – Lâmina própria intestinal;

mLN – Linfonodo mesentérico;

IgG – Imunoglobulina do tipo G;

IgA – Imunoglobulina do tipo A;

CD – Células dendríticas;

CD71 – Células dendrítica do tipo 71;

CD103+ – Células dendríticas do tipo 103 selecionadas positivamente;

CD4 – Células dendríticas do tipo 4;

MHC II – Complexo de histocompatibilidade do tipo II;

TGF- $\beta$  – fator de crescimento e expansão celular do tipo  $\beta$ ;

RA – Ácido retinóico;

iTREGFOXP3+ – Células T regulatórias induzidos do tipo FOXP3 selecionadas positivamente;

IL-10 – Interleucina do tipo 10;

IL-27 – Interleucina do tipo 27

CD11b+ – Células dendríticas do tipo 11b selecionadas positivamente;

TR1 – Células T regulatórias do tipo 1;

CD3+ - Células dendríticas do tipo 3 selecionadas positivamente;

INF-lambda – Interferon lambda;

TH3 – Células auxiliares do tipo 3;

nTREGFOXP3+ – Célular T regulatórias naturais do tipo FOXP3 selecionadas positivamente;

“Naïve”CD3FOXP3- Células dendríticas “ingênuas” do tipo FOXP3 selecionadas negativamente;

CCR7 – Proteína integral de membrana codificada pelo gene CCR7;

IL-2 – Interleucina do tipo 2;

FcRn – Receptor de membrana neonatal do tipo Fc;

$\alpha 4\beta 7$  integrina – Elemento protein integral de membrane do tipo  $\alpha 4\beta 7$ ;

MadCAM-1 – Molécula de adesina musosal vascular de adesão celular do tipo 1;

CCL25 – Citosina do tipo 25;

CD11c+ – Células dendríticas do tipo 11c selecionadas positivamente;

pCDs – Células dendríticas do tipo p;

CD11b+ – Célula dendríticas do tipo 11b selecionadas positivamente;

TH1 – Células auxiliares do tipo 1;

TH2 – Células auxiliares do tipo 2;

TH17 – Células auxiliares do tipo 17;

CD103- – Células dendríticas do tipo 103 selecionadas negativamente;

CD11b- – Células dendríticas do tipo 11b selecionadas negativamente;

RALDH2 – Enzima Retinaldeído desidrogenase tipo 2;

p38 $\alpha$  – Anticorpo MAP quinase;

CD4<sup>+</sup>CD49b<sup>+</sup>LAG-3<sup>+</sup> –

FOXP3 – Proteína responsável pela diferenciação das células iTREGFOXP3;

LAG-3 – Gene de ativação linfocitária.

# Lista de símbolos

$V$  – Conjunto de vértices;

$V^{(2)}$  – Conjunto de pares ordenados especificados por vértices;

$A$  – Conjunto de arestas;

$G$  – Grafo genérico;

$v_i$  – Vértice genético  $i$ ;

$a_i$  – Aresta genérica  $i$ ;

$\emptyset$  – Conjunto vazio;

§ – Parágrafo;

$N(v)$  – Número de vizinhos do vértice  $v$ ;

$\delta(v)$  – Grau do vértice  $v$ ;

$A^*$  – Conjunto de arestas de um subgrafo;

$V^*$  – Conjunto de vértices de um subgrafo;

$A$  – Matriz de adjacência genérica;

$A[i, j]$  – Elemento da matriz de adjacência  $A$  com entradas  $i$  e  $j$ ;

$M$  – Matriz de incidência genérica;

$M[i, j]$  – Elemento da matriz incidência  $M$  com entradas  $i$  e  $j$ ;

$\forall$  – Para todo;

$\exists$  – Existe;

| – Tal que;

: – Tal que;

$\in$  – Pertence a;

$\subset$  – Subconjunto de;

$\subseteq$  – Subconjunto de ou e igual a;

$seq$  – Sequência genérica de valores;

$seq.ord$  – Sequência ordenada de valores;

$R$  – Conjunto de arcos;

$r_i$  – Arco genérico  $i$ ;

$D$  – Grafo direcionado ou dígrafo;

$k_i$  – Notação alternativa para o grau de um vértice  $i$ ;

$\langle k \rangle$  – Grau médio;

$p$  – Probabilidade;  
 $q$  – Probabilidade;  
 $Pr$  - Probabilidade;  
 $C_N^k$  – Combinação de  $k$  graus tomados  $N$  a  $N$ ;  
 $z$  – Parâmetro de sintonia;  
 $\ell_G$  – Distância de caminho médio de uma rede;  
 $N_i$  – Número observado de vizinhos de um vértice  $i$ ;  
 $K_i$  – Número possível de vizinhos de um vértice  $i$ ;  
 $d(v_i, v_j)$  – Menor distância entre os vértices  $i$  e  $j$ ;  
 $C_i$  – Coeficiente de aglomeração do vértice  $i$ ;  
 $\prod(k_i)$  - Probabilidade em função do grau do vértice  $i$ ;  
 $\ell_{max}$  - Diâmetro da rede;  
 $\langle C_i \rangle$  – Coeficiente de aglomeração médio da rede;  
 $g_l$  – Centralidade de intermediação do vértice  $l$ ;  
 $\Pi$  – Distribuição estacionária;  
 $h_{uv}$  - Probabilidade de alcançar  $v$  de  $u$  em  $t$  passos;  
 $G_c$  – Tamanho da componente gigante de um grafo  $G$ ;  
 $r$  – Assortatividade de grau;  
 $\bar{C}(k)$  - Coeficiente de aglomeração médio como função do grau;  
S.C. – Subgrafo componente;  
P.C. – Partição componente;  
 $p(k)$  - Distribuição de graus;  
 $k_{i\ out}$  - Grau de saída do vértice  $i$ ;  
 $N(t)$  - Número total de caminhantes na rede como função do tempo;  
 $W(G; t)$  – Vetor de posições do grafo  $G$  no instante  $t$ ;  
 $w_i(t)$  – Posição do  $i$ -ésimo caminhante na rede no instante  $t$ ;  
 $S(G; t)$  - Vetor de frequências absolutas de caminhantes no grafo  $G$  no instante  $t$ ;  
 $\sigma_i(t)$  - Número de caminhantes no vértice  $i$  no instante  $t$ ;  
 $F(G; t)$  – Vetor de fluxo ou vetor de frequências relativas de caminhantes no grafo  $G$  no instante  $t$ ;  
 $f_i(t)$  – Fluxo ou valor relativo de caminhantes no vértice  $i$  no instante  $t$ ;  
 $\Pi(G)$  - Distribuição estacionária de um grafo  $G$  num tempo assintótico ou vetor de fluxos em um tempo assintótico de um grafo  $G$ ;

$\Delta f_i$  - Diferença entre os valores de fluxo entre dois grafos semelhantes;

$p_i$  - Função erro relativo de fluxo;

$P(G; G'; t \rightarrow \infty)$  - Erro relativo médio entre os grafos  $G$  e  $G'$  num tempo assintoticamente grande;

$T(G)$  - Matriz de transição de um grafo  $G$  qualquer;

$\{\lambda_i\}$  - Conjunto de autovalores;

$I$  - Matriz identidade;

$\bar{F}(G; t; L)$  - Vetor de fluxo médio no grafo  $G$  no instante  $t$  com  $L$  repetições;

$\bar{P}(G; G'; t)$  - Erro relativo médio estatístico entre os grafos  $G$  e  $G'$  num tempo computacionalmente escolhido;

# Sumário

## Lista de Figuras

## Lista de Tabelas

|  |      |
|--|------|
| <b>Capítulo 1. Introdução</b> .....                                    | p.16 |
| <b>Capítulo 2. Teoria de Grafos e Análise de Redes Complexas</b> ..... | p.20 |
| 2.1 Teoria de grafos.....  | p.20 |
| 2.1.1 Grafo.....   | p.21 |
| 2.1.2 Vizinhos de um vértice.....                                      | p.23 |
| 2.1.3 Grau de um vértice.....  | p.23 |
| 2.1.4 Subgrafo de um grafo.....  | p.25 |
| 2.1.5 Representação de um grafo: matriz de adjacência.....             | p.25 |
| 2.1.6 Isomorfismo entre grafos.....                                    | p.27 |
| 2.1.7 Conectividade em um grafo.....                                   | p.29 |
| 2.1.8 Grafos direcionados ou dígrafos.....                             | p.30 |
| 2.2 Redes complexas: análise e caracterização.....                     | p.33 |
| 2.2.1 Considerações gerais sobre redes complexas.....                  | p.34 |
| 2.2.2 Grafos e redes complexas: histórico e desenvolvimento.....       | p.35 |
| 2.2.3 Teoria de redes randômicas.....                                  | p.38 |
| 2.2.4 Subgrafos em redes randômicas.....                               | p.40 |
| 2.2.5 Evolução de redes randômicas.....                                | p.42 |
| 2.2.6 Teoria de grafos “small world”.....                              | p.43 |
| 2.2.7 Teoria de redes livre de escala.....                             | p.45 |
| 2.2.8 Caracterização de redes complexas.....                           | p.47 |
| 2.3 Considerações finais do capítulo.....                              | p.50 |
| <b>Capítulo 3. Caminhada Aleatória em Redes Complexas</b> .....        | p.51 |
| 3.1 Caminhada aleatória.....   | p.51 |
| 3.1.1 Caminhada aleatória: histórico e desenvolvimento.....            | p.51 |
| 3.1.2 Caminhada aleatória em uma dimensão.....                         | p.52 |
| 3.1.3 Caminhada aleatória em duas dimensões.....                       | p.54 |
| 3.1.4 Caminhada aleatória em redes complexas.....                      | p.55 |
| 3.1.5 Teorema fundamental da cadeia de Markov.....                     | p.58 |



|   |              |
|---|--------------|
| 3.2 Considerações finais do capítulo.....   | p.59         |
| <b>Capítulo 4. A Rede Mitológica da Odisseia de Homero.....</b>                     | <b>p.60</b>  |
| 4.1 Mitologia comparativa e redes mitológicas.....                                  | p.60         |
| 4.1.1 Contextualização do uso de redes complexas em mitologia comparativa.....      | p.60         |
| 4.1.2 Descrição literária do poema da Odisseia.....                                 | p.62         |
| 4.2 Análise da rede mitológica da Odisseia.....                                     | p.63         |
| 4.2.1 Coleta de dados textuais.....   | p.63         |
| 4.2.2 Análise indutiva.....   | p.65         |
| 4.2.3 Análise dedutiva.....   | p.72         |
| 4.3 Considerações finais do capítulo.....   | p.77         |
| <b>Capítulo 5. A Tolerância Oral como Fenômeno de Redes Complexas.....</b>          | <b>p.79</b>  |
| 5.1 Imunologia e tolerância oral.....   | p.79         |
| 5.1.1 Conceitos gerais acerca da resposta imunológica adaptativa.....               | p.79         |
| 5.1.2 Tolerância oral e complexidade.....   | p.81         |
| 5.2 Tolerância oral e redes complexas.....  | p.81         |
| 5.2.1 Construção do modelo de rede.....   | p.81         |
| 5.3 Caminhada aleatória como modelo de difusão de estímulos em redes complexas..... | p.84         |
| 5.3.1 Modelo dinâmico estocástico.....  | p.84         |
| 5.3.2 Método de análise provido pelo modelo estocástico.....                        | p.87         |
| 5.3.3 Aplicação do método no contexto da tolerância oral.....                       | p.89         |
| 5.3.4 Considerações algébricas acerca do modelo estocástico.....                    | p.89         |
| 5.3.5 Método algorítmico de implementação do modelo.....                            | p.91         |
| 5.4 Resultados e discussão.....   | p.92         |
| 5.4.1 Resultados.....   | p.92         |
| 5.4.2 Discussão: síntese biológica acerca dos resultados do modelo.....             | p.94         |
| 5.5 Considerações finais do capítulo.....   | p.98         |
| <b>Capítulo 6. Conclusões e Trabalhos Futuros.....</b>                              | <b>p.99</b>  |
| <b>Referências bibliográficas.....</b>  | <b>p.103</b> |
| <b>Anexo A.....</b>   | <b>p.109</b> |

# CAPÍTULO 1

## Introdução

O reducionismo tem sido a interpretação científica dominante acerca dos objetos e fenômenos naturais desde o século XVII. Sua notoriedade histórica inicia com Galileu Galilei, mas é com René Descartes, em sua obra: *Discours de la méthode* (Discurso do Método) [1] que a mesma se torna consagrada. Nesta obra Descartes discursa sobre o método científico de apreensão da realidade, dividindo-a em quatro processos fundamentais:

- “1. Coletar as informações cuidadosamente, examinando sua racionalidade e justificação. Verificar a veracidade e boa procedência do investigado, obtendo apenas o indubitável.”
- “2. Analisar, dividir ou quebrar em partes o objeto de estudo tanto quanto for possível;”
- “3. Gerar a síntese ou elaborar progressivamente as conclusões mais abrangentes e ordenadas a partir dos objetos mais simples e fáceis até os mais complexos e difíceis;”
- “4. Enumerar e revisar minuciosamente as conclusões, garantindo que nada seja omitido e que a coerência geral exista.”

Estes quatro processos enfatizam o escrutínio de todas as informações acerca do sistema físico, reforçando a base empiricista de prova entre hipóteses possíveis por meio dos dados estudados separadamente. Todo reducionismo se estabelece com a convenção de que o sistema pode ser reduzido à partes menos complexas que o todo, estudando-as separadamente, e então interpretando o comportamento total do sistema.

Desde Newton, Descartes, Leibniz e outros pensadores da época, este tem sido o único método reconhecidamente científico de apreensão, análise e síntese acerca das coisas naturais. As explicações geradas por esse processo com fins científicos foi o único a ser aplicado até o início do século XX, onde toda a interpretação fenomenológica da física foi construída. A posterioridade deste pensamento veio com Albert Michelson [2], que dissertou em 1894 a seguinte questão:

“... parece provável que a maior parte dos princípios fundamentais da física já foram estabelecidos, e que seu avanço neste momento, só depende da aplicação rigorosa dos mesmos aos fenômenos mais complexos à medida que se tornam apreensíveis.”

Inesperadamente, mal podia prever Michelson que, após trinta anos iria surgir a mecânica quântica e a relatividade que juntamente com o desenvolvimento das equações de movimento não lineares (caos), colocariam em xeque seu “sonho reducionista”. De uma forma geral, o século XX foi marcadamente caracterizado por essa destruição da ambição de descrever a realidade a partir de princípios puramente reducionistas, ou seja, o “sonho reducionista” estava desfeito. No entanto, era evidente o sucesso da física até então fundamentada, pois a maioria dos problemas em escalas muito pequenas ou muito grandes são bem descritas. Logo, o ponto de “quebra” das generalizações obtidas pela física reducionista surgia à medida que os fenômenos de escalas médias eram analisados, ou seja, fenômenos na escala humana [3].

Muitos destes fenômenos foram inapreensíveis para o paradigma da física daquela época, como: a imprevisibilidade do tempo e do clima, o comportamento econômico, as preferências políticas, sociais e/ou culturais. E, acima de tudo, o que mais impressionava os cientistas e sua visão das coisas mínimas, era o comportamento complexo de sistemas vivos [4]. Desta frustração analítica surge a famosa frase: “o todo é mais do que a soma das partes”; ao qual vem se formalizando nas novas ciências: dinâmica caótica, biologia sistêmica, dinâmica econômica e a teoria de redes. Todas essas aproximações estão muito além da abordagem reducionista dos fenômenos, se aproximando a um pensamento mais holista.

Holos, do grego, significa inteiro, completo ou total. O holos de um sistema físico indica a irreducibilidade de algumas de suas propriedades, aquelas características globais que não podem ser explicadas apenas pela soma das suas componentes. A priori, constitui uma visão “de cima para baixo”, processo inverso ao reducionismo, “de baixo para cima”. A Aristóteles atribui-se o pioneirismo desta concepção de completude dos objetos físicos, indicado vastamente em seu Tratado de Metafísica, mas sintetizado na forma: “*O todo é maior do que a simples soma das suas partes*” [5].

Passados quase 22 séculos, foi só a partir da segunda metade do século XX que os filósofos e estudiosos retornaram a visão emergentista, gerando uma quebra de paradigma científico: reducionismo versus holismo. Pouco a pouco, o reducionismo dava margem para a interpretação de fenômenos concebidos pelo novo paradigma, ao qual não podia ser atribuído a uma única disciplina, mas a um entendimento interdisciplinar que ainda não estava bem

fundamentado [6]. Muitas destas tentativas interdisciplinares geraram áreas como cibernética, sinérgica, ciência de sistemas, ou, mais recentemente a teoria de sistemas complexos. Com este mesmo objetivo, um grupo de vinte e quatro cientistas e matemáticos inicialmente se reuniram, em 1984, para discutir esse assunto interdisciplinar de modo a formar um corpo conceitual aproximadamente estruturado.

O marco deste evento ocorreu no alto do deserto de Santa Fé, Novo México. Logo, com o tempo, o Instituto de Santa Fé foi criado e escolhido como o centro de pesquisa sobre a nova ciência dos sistemas complexos. Este local é reconhecido por receber pessoas de diversas formações, países e culturas que possuíam apenas um aspecto em comum: responder às mesmas perguntas, mas com metodologias diferentes [7].

O meio pelo qual os pesquisadores da época esperavam atingir esse objetivo consistia na análise de grandes pacotes de dados, “Big Data”, os quais estavam associados aos sistemas observados na natureza (*e. g.*, ecossistemas, redes de energia, redes de comunicação, redes sociais, etc.) que possuiria caráter complexo e interativo, ou seja, um grande número de ligações entre as componentes do sistema [8]. A partir deste princípio, então, promovia-se uma unidade do conhecimento acerca dos sistemas, agora nomeados complexos, onde a teoria poderia se organizar e se tornar a pedra fundamental na construção do pensamento sistêmico.

Considerando todo esse “background” histórico do desenvolvimento do pensamento sistêmico acerca dos fenômenos naturais, há um crescente interesse em responder às questões mais elementares sobre a complexidade, adaptação e emergência de redes complexas reais; como as conexões de regiões no cérebro, a dependência mútua na interação inseto-planta, a dependência topológica do sistema imune para uma resposta infecciosa, as relações intercelulares, a economia global ou na evolução biológica. O que é notável nestes sistemas é sua natureza inter-relacionada quando estudamos suas partes. Como resposta a estudos deste tipo, surgiu a teoria de redes complexas visando modelar a interrelação e complexidade interacional entre as partes. Para tanto, lança-se mão de alguns artifícios matemáticos encontrados na teoria de grafos e de redes complexas.

Neste contexto o estudo de redes sociais tem fomentado o desenvolvimento da teoria de redes complexas que, em última análise, estuda sistemas complexos [33, 35, 36, 38]. O interesse no estudo de redes sociais aumentou nos últimos anos, dado os interesses políticos, econômicos e científicos desta linha de pesquisa. Com isso, este trabalho é composto por um estudo de caso acerca da concepção da rede social mitológica da Odisseia, coletada por meio da mineração textual [52] e analisada pelo escopo da teoria de redes complexas. Analogamente, mas em um contexto diferente, o modelo de redes complexas tem sido de

grande valia para a análise das relações moleculares de sistemas biológicos de múltiplas componentes. Um destes sistemas constitui o tecido linfóide associado ao intestino, que é responsável pelo fenômeno da tolerância oral [58, 59]. Este fenômeno, experimentalmente, se torna quase inapreensível ao pesquisador, pois o seu funcionamento depende de relações complexas de suas partes [59-73]. Dadas estas complicações, é pretendido neste trabalho o estudo deste fenômeno por vias da teoria de redes complexas e caminhada aleatória como gerador de um novo conceito de difusão de estímulos na rede.

Com isso, o objetivo geral deste trabalho consiste em utilizar a chave conceitual da emergência como justificativa do uso da teoria de redes complexas no estudo de dois sistemas reais diferentes: uma rede social mitológica e uma rede imunológica.

A organização textual deste manuscrito segue da seguinte maneira: no capítulo 2 contém um apanhado geral sobre o essencial da teoria de grafos como fundamento para o estudo de redes complexas. No capítulo 3 são estudados os processos markovianos e a caminhada aleatória em redes complexas. No capítulo 4 será uma aplicação dos conceitos explorados no capítulo 2, onde é pretendido o estudo da mitologia comparativa da rede social da Odisseia de Homero. Finalmente, no capítulo 5, será explorada a caminhada aleatória em redes complexas como modelo de transmissão de informação para o estudo de um fenômeno imunológico: a tolerância oral.

## CAPÍTULO 2

# Teoria de Grafos e Análise de Redes Complexas

Neste capítulo pretende-se fazer uma abordagem simples da teoria de grafos de forma a gerar condições para o estudo de redes complexas, e então, redes reais. O conceito de emergência é a justificativa do uso do objeto grafo como o modelo mais adequado de descrição de fenômenos complexos e/ou emergentes. Em defesa desta postura, estudam-se aqui as formas de análise de redes e os modelos atuais de evolução de redes. Deste capítulo pretende-se fundamentalmente a identificação de quantificadores que justifiquem a comparação entre redes, de modo a gerar uma nova forma de análise contemplando a posição emergentista da realidade.

## 2.1 Teoria de Grafos

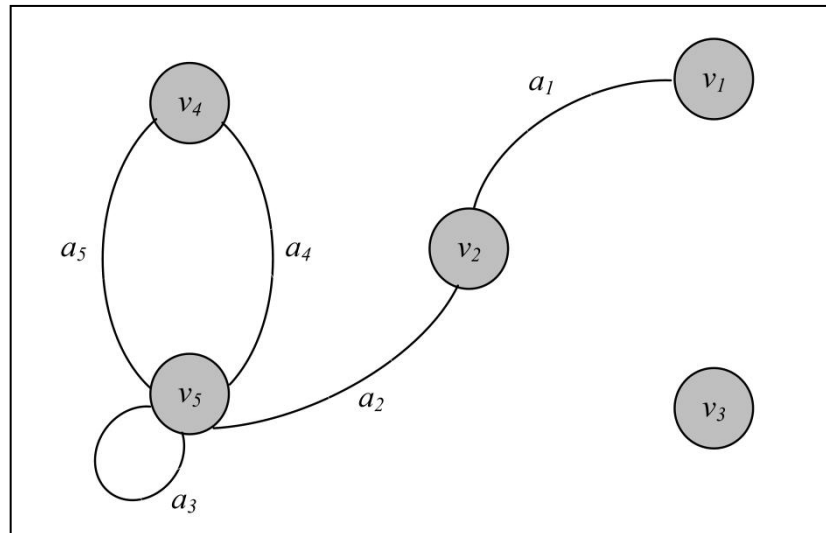
Nesta seção são abordados os conceitos gerais da teoria de grafos como elementos que fundamentam a análise de redes complexas, que é um dos principais objetos de estudo deste manuscrito. A teoria de grafos, em termos grosseiros, é o ramo da matemática que trata do objeto matemático denominado grafo, que são estruturas compostas por partes e conexões entre estas partes. Este capítulo será uma revisão, compilação e síntese das seguintes bibliografias: Van Mieghem [9], Van Steen [10] e Rouhonen [11].

### 2.1.1 Grafo

Pictoricamente, um grafo é composto por vértices e arestas que conectam vértices.

Exemplo:

Figura 2.1 Exemplo de um grafo.



Fonte: autor

Formalmente, para qualquer conjunto de vértices  $V$ , denotaremos por  $V^{(2)}$  como o conjunto de todos os pares não ordenados de elementos de  $V$ . Se  $V$  tem  $n$  elementos então  $V^{(2)}$  tem  $\binom{n}{2} := \frac{n(n-1)}{2}$  elementos. Os elementos de  $V^{(2)}$  serão identificados com os subconjuntos de  $V$  que têm cardinalidade 2, constituindo o conjunto de arestas  $A$ . Assim, cada elemento de  $V^{(2)}$  terá a forma  $(v, w)$ , sendo  $v$  e  $w$  dois vértices quaisquer de  $V$ . Com isso, é possível escrever:

**Definição 2.1.** Um grafo é um par  $(V, A)$  em que  $V$  é um conjunto arbitrário e  $A$  é um subconjunto de  $V^{(2)}$ , de forma que os elementos de  $V$  são chamados vértices e os de  $A$  são chamados arestas.

É possível descrever o grafo da Figura 1 da seguinte maneira; um par ordenado  $G = (V, A)$ , de forma que  $V = \{v_1, \dots, v_5\}$  é o conjunto de vértices e  $A = \{(v_1, v_2), (v_2, v_5), (v_5, v_5), (v_5, v_4), (v_4, v_5)\}$  o conjunto de arestas. É comum, por

simplicidade, rotular todas as arestas em termos de um grupo estanco, onde cada elemento deste grupo é representado por uma letra ou número, de forma que

$$A = \{(v_1, v_2), (v_2, v_5), (v_5, v_5), (v_5, v_4), (v_4, v_5)\} \therefore$$

$$A = \{(v_1, v_2) = a_1, (v_2, v_5) = a_2, (v_5, v_5) = a_3, (v_5, v_4) = a_4, (v_4, v_5) = a_5\} \therefore$$

$$A = \{a_1, a_2, a_3, a_4, a_5\}.$$

Uma importante ressalva: os pares  $(v_5, v_4)$  e  $(v_4, v_5)$  são algebricamente equivalentes, em outras palavras, para este caso os pares não precisam ser necessariamente ordenados. Isso é particularmente verdadeiro para grafos não direcionados, em que as arestas não possuem direcionalidade. Nas seções posteriores será desenvolvida essa terminologia.

À medida que definimos grafo por meio da definição 1, algumas consequências lógicas estão quase sempre associadas a terminologias usuais da teoria de grafos. A seguir, uma lista das mais importantes terminologias e seus exemplos:

**1** – Dois vértices quaisquer  $v$  e  $u$  são vértices de extremidade da aresta  $(v, u)$ , quando constituem a porção distal da aresta  $(v, u)$ . Todos os elementos de  $V$  correspondem a este tipo de vértice;

**2** – Arestas que têm os mesmos vértices são denominadas arestas adjacentes. As arestas  $a_1$  e  $a_2$  são arestas adjacentes;

**3** – Uma aresta na forma  $(v, v)$  é uma alça, ou do inglês, “loop”. A aresta  $a_3$  é uma aresta do tipo alça;

**4** – Um grafo será denominado grafo simples se nele não houver alças ou arestas paralelas;

**5** – Um grafo sem arestas (*i. e.*,  $A = \{\emptyset\}$ ) é denominado grafo vazio;

**6** – Um grafo sem vértices (*i. e.*,  $V = \{\emptyset\}$  e  $A = \{\emptyset\}$ ) é denominado grafo nulo;

**7** – Um grafo constituído de apenas um vértice é denominado grafo trivial;

**8** – Duas arestas quaisquer são denominadas arestas paralelas se compartilham um vértice. As arestas  $a_5$  e  $a_4$  são arestas paralelas;

**9** – Analogamente a §8, quaisquer dois vértices  $u$  e  $v$  serão vértices adjacentes se houver uma aresta que os conectem,  $(u, v)$ ;



**10** – Denominamos vértice suspenso qualquer vértice que possuir grau 1 (*i. e.*, número de arestas incidentes ao vértice), ou pictoricamente, o vértice que estiver associado a apenas uma aresta. O vértice  $v_1$  é um vértice suspenso;

**11** – Como consequência de §8, quando uma aresta tem como uma de suas extremidades um vértice suspenso, esta é denominada aresta suspensa. A aresta  $a_1$  é uma aresta suspensa;

**12** – Um vértice que possui grau igual à zero, ou seja, não está associada a nenhuma aresta, é denominado vértice isolado. O vértice  $v_3$  é um vértice isolado;

**13** – Um grafo simples com  $n$  vértices que sejam todos vértices adjacentes, é denominado grafo completo;

### 2.1.2 Vizinhos de um vértice

Em um sentido prático, é conveniente introduzir o conceito de vizinhos de um vértice. É possível falar no mesmo sentido em se tratando do conceito de vértices adjacentes, mas como este é um dos elementos centrais da teoria de grafos, escrevamos:

**Definição 2.2.** *Para um grafo qualquer  $G$  e vértice  $v \in A(G)$ , o conjunto de vizinhos  $N(v)$  de  $v$  é o conjunto de vértices adjacentes a  $v$ , que satisfaz*

$$N(v) \stackrel{\text{def}}{=} \{w \in A(G) \mid v \neq w, \exists a \in A(G) : a = (w, v)\}. \quad (2.1)$$

A notação acima pode ser interpretada de forma simples: seja  $N(v)$  o número de vizinhos de um vértice qualquer  $v$  e se houver uma aresta que conecta este vértice a um vértice  $w$  qualquer, este será vizinho de  $v$ .

### 2.1.3 Grau de um vértice

À medida que se definem os vizinhos de um vértice, fica evidente uma propriedade estrutural: o número de arestas incidente a ele. A esta propriedade, em teoria de grafos, denominamos grau do vértice. Em termos mais precisos:

**Definição 2.3.** O número de arestas incidentes a um vértice qualquer  $v$  é denominado grau de  $v$ , denotado  $\delta(v)$ . Arestas to tipo alça são contadas duas vezes na valoração do grau.

Quando se tem um grafo  $G$  e se contabilizam os graus de todos os seus vértices, inevitavelmente encontra-se o seguinte teorema:

**Teorema 2.1.** Para todo grafo  $G$ , a soma dos graus dos vértices é igual a duas vezes o número total de arestas, ou seja,

$$\sum_{v \in V(G)} \delta(v) = 2|A(G)|. \quad (2.2)$$

**Prova.** Quando se contam as arestas de um grafo  $G$ , contabilizando-os em cada vértice  $v$  de  $G$ , em verdade está se contabilizando exatamente duas vezes cada aresta. Logo

$$\sum_{v \in V(G)} \delta(v) = \delta(v_1) + \delta(v_2) + \delta(v_3) + \cdots \delta(v_n). \quad (2.3)$$

mas como

$$\delta(v_i) = |\{a_{ij} = (v_i, u_j)\}|, \text{ onde } j = 1, 2, 3, \dots, n; \text{ são os vizinhos de } i. \quad (2.4)$$

Estabelecendo que:

$$a_{ij} = a_{ji}. \quad (2.5)$$

Isso mostra que quando contamos os elementos de  $\delta(v_i)$  e  $\delta(v_j)$ , sendo que exista algum  $a_{ij}$ , invariavelmente existira um  $a_{ji}$ , mostrando que a mesma aresta é contabilizada duas vezes quando somamos todos os graus dos vértices de um grafo.

### 2.1.4 Subgrafo de um grafo

Outro conceito importante e usual em teoria de grafos é de subgrafo. De uma forma muito geral, um subgrafo, estruturalmente, também é um grafo. Sendo assim, se  $H$  é um grafo e subgrafo do grafo  $G$ , então  $H$  deve consistir de um subconjunto de arestas e vértices de  $G$ . Isso deve ser satisfeito de modo que para toda extremidade de uma aresta em  $H$  deva haver um vértice em  $H$ . Logo:

**Definição 2.4.** *Um grafo  $H$  é um subgrafo de  $G$  se  $V(H) \subseteq V(G)$  e  $A(H) \subseteq A(G)$  de modo que para todo  $a \in A(H)$  com  $a = (u, v)$ , de modo que  $u, v \in V(H)$ . Quando  $H$  é subgrafo de  $G$ , denota-se  $H \subseteq G$ .*

No estudo de subgrafos, é conveniente considerar subgrafos definidos por um subconjunto de vértices que pertençam ao grafo. Quando isso se efetua na identificação de um subgrafo, é denominado subgrafo induzido. Ou seja, toma-se um subconjunto de vértices  $V^*$  e todas as suas arestas que os conectam. Formalmente:

**Definição 2.5.** *Considere um grafo  $G$  e um subconjunto dos seus vértices  $V^* \subseteq V(G)$ . O subgrafo induzido por  $V^*$  é constituído pelos vértices  $V^*$  e as arestas  $A^*$  definidas por*

$$A^* \stackrel{\text{def}}{=} \{a \in A(G) \mid a = (u, v) \text{ com } u, v \in V^*\}. \quad (2.6)$$

*Analogamente, se  $A^* \in A(G)$ , o subgrafo induzido por  $A^*$  é composto pelo conjunto de arestas  $A^*$  e o conjunto de vértices  $V^*$ , dado por*

$$V^* \stackrel{\text{def}}{=} \{u, v \in V(G) \mid \exists a \in A^*: a = (u, v)\}. \quad (2.7)$$

*O subgrafo induzido por  $V^*$  ou  $A^*$  é denotado por  $G[V^*]$  ou  $G[A^*]$ , respectivamente.*

### 2.1.5 Representação de um grafo: matriz de adjacência

Um grafo pode ser representado de várias maneiras, tanto pictoricamente quanto por elementos algébricos. Porém, há um apelo à facilidade no uso da matriz de adjacência de um

grafo. A matriz de adjacência  $\mathbf{A}[i, j]$  nada mais é do que um arranjo com  $n$  linhas e  $n$  colunas, que denota o número de arestas que conecta o vértice  $v_i$  ao vértice  $v_j$ . Utilizando a figura 2.1 como modelo, teríamos a seguinte matriz de adjacência:

Tabela 2.1 Matriz de adjacência do grafo da Figura 2.1.

|       | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ |
|-------|-------|-------|-------|-------|-------|
| $v_1$ | 0     | 1     | 0     | 0     | 0     |
| $v_2$ | 1     | 0     | 0     | 0     | 0     |
| $v_3$ | 0     | 0     | 0     | 0     | 0     |
| $v_4$ | 0     | 0     | 0     | 0     | 2     |
| $v_5$ | 0     | 0     | 0     | 2     | 1     |

Fonte: autor.

Algumas propriedades da matriz de adjacência podem ser identificadas:

1. Uma matriz de adjacência é simétrica, ou seja, para todo  $i, j$ ,  $\mathbf{A}[i, j] = \mathbf{A}[j, i]$ . Esta propriedade reflete o fato de que uma aresta é representada por um par não ordenado de vértices de modo que  $a = (v_i, v_j) = (v_j, v_i)$ ;
2. Um grafo  $G$  será simples se e somente se para todo  $i, j$ ,  $\mathbf{A}[i, j] \leq 1$  e  $\mathbf{A}[i, j] = 0$ . Em outras palavras, é o caso de grafos sem alças nem arestas duplas: cada par de vértices pode compartilhar apenas uma aresta;
3. A soma dos valores de uma linha  $i$  será igual ao grau do vértice  $v_i$ , isto é,  $\delta(v_i) = \sum_{j=1}^n \mathbf{A}[i, j]$ .

Uma forma alternativa de representar um grafo é por meio da matriz de incidência. Uma matriz de incidência  $\mathbf{M}$  de um grafo  $G$  consiste em  $n$  linhas e  $m$  colunas de modo que  $\mathbf{M}[i, j]$  contabiliza o número de vezes que uma aresta  $a_j$  é incidente ao vértice  $v_i$ . Note que  $\mathbf{M}[i, j]$  será tanto 0, 1 ou 2, a medida que uma aresta, respectivamente, não ser incidente ao vértice  $v_i$ , ou possuir uma de suas extremidades associados ao  $v_i$  ou ser uma alça de  $v_i$ . A seguinte tabela constitui a matriz de incidência do grafo representado na figura 2.1:

Tabela 2.2 Matriz de incidência do grafo da Figura 2.1.

|       | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|-------|-------|-------|-------|-------|-------|
| $v_1$ | 1     | 0     | 0     | 0     | 0     |
| $v_2$ | 1     | 1     | 0     | 0     | 0     |
| $v_3$ | 0     | 0     | 0     | 0     | 0     |
| $v_4$ | 0     | 0     | 0     | 1     | 1     |
| $v_5$ | 0     | 0     | 2     | 1     | 1     |

Fonte: autor.

Algumas propriedades da matriz de incidência são de citação relevante:

1. Um grafo  $G$  não possui alças se e somente se para todo  $i, j$   $M[i, j] \leq 1$ ;
2. A soma de todos os valores em uma linha  $i$  deve ser igual ao grau do vértice  $v_i$ .  
Em termos matemáticos, é expresso por  $\forall i: \delta(v_i) = \sum_{j=1}^m M[i, j]$ ;
3. Como cada aresta tem exatamente duas extremidades, para todo  $i$ ,  
 $\sum_{j=1}^m M[i, j] = 2$ .

Apesar de algumas facilidades analíticas proporcionadas pelas matrizes de adjacência e de incidência, sua expressão se torna inconveniente no tratamento de dados para grafos muito grandes. De modo que a matriz de adjacência cresce com o quadrado do número de vértices e a matriz de incidência cresce na proporção  $n \times m$ . Dadas estas peculiaridades acerca da representação matricial de um grafo, introduz-se o conceito de lista de arestas. É basicamente a lista de arestas de um grafo  $G$ , onde cada aresta é especificada pelos vértices em que é incidente. Nesta notação, o tamanho da lista de arestas cresce linearmente com o número de arestas, ou seja, é uma formação mais compacta na representação de grafos. Por exemplo, se fossemos representar o grafo da Figura 3.1 em termos de sua lista de arestas, obteríamos:

$$\{a_1 = (v_1, v_2), a_2 = (v_2, v_3), a_3 = (v_5, v_5), a_4 = (v_4, v_5), a_5 = (v_4, v_5)\}. \quad (2.8)$$

Neste caso, a representação total de um grafo deve vir acompanhada, também, por uma lista de vértices. Pois há a possibilidade de um grafo ter vértices isolados de modo que apenas a lista de arestas seria insuficiente para descrever o grafo totalmente.

### 2.1.6 Isomorfismo entre grafos

É de importância notar que, de todas as representações de grafos citadas até agora, nenhuma dá pistas da representação pictórica exata do grafo. Embora qualquer configuração seja possível pictoricamente para um grafo qualquer ser representado por meio de matriz de adjacência, matriz de incidência e lista de arestas, o objeto, grafo, continua o mesmo. Sendo assim, essa similaridade entre a representação do grafo e sua forma pictórica pode ser formalizada.

**Definição 2.6.** *Considere dois grafos  $G = (V, A)$  e  $H = (V', A')$ . Sejam  $G$  e  $H$  isomórficos, então haverá um mapeamento um a um do tipo  $\phi: V \rightarrow V'$  de modo que para toda aresta  $a \in A$  sendo  $a = (u, v)$ , haverá um único vértice  $a' \in A'$  que corresponde a  $a' = (\phi(u), \phi(v))$ .*

Isso significa que, dados dois grafos que são isomórficos, de modo que seus vértices sejam conhecidos, necessariamente o mesmo número de arestas deve conectá-los de modo a respeitar certa distribuição de grau. Para tanto é necessário conceituar distribuição de grau de um grafo.

A distribuição de grau de um grafo é simplesmente a lista dos graus de todos os vértices do grafo. É comumente uma lista com valores decrescente na forma de uma sequência ordenada. Por exemplo, a distribuição de graus associada ao grafo da figura 2.1 é dada por:

$$\begin{aligned} & seq[\delta(v_1), \delta(v_2), \delta(v_3), \delta(v_4), \delta(v_5)] \therefore \\ & seq[1, 2, 0, 2, 4]; \text{ordenando:} \\ & seq.ord[4, 2, 2, 1, 0]. \end{aligned}$$

Em que *seq* é uma sequência de valores não ordenados, e *seq.ord* é uma sequência de valores ordenados, no caso acima ordenados decrescentemente. Algumas facilidades são proporcionadas no estudo de grafo utilizando a distribuição de graus. Por exemplo, quando todos os vértices possuem mesmo grau, isso significa que se trata de um grafo regular.

### 2.1.7 Conectividade em um grafo

Dos elementos de grafos que se tratou até agora, para cada vértice  $v$  assumiu-se que pudesse ser acessado por qualquer outro vértice  $w$ , se o grafo não possuir vértices isolados, de modo que haja uma cadeia de vértices adjacentes entre  $v$  e  $w$ . Essa intuição permite a seguinte definição de conectividade:

**Definição 2.7.** *Considere um grafo  $G$  e uma caminhada do tipo  $(v_0, v_k)$  de modo que seja uma sequência alternante  $[v_0, a_1, v_1, a_2, \dots, v_{k-1}, a_k, v_k]$  de vértices e arestas de  $G$  de modo que  $a_i = (v_{i-1}, v_i)$ . Em uma caminhada fechada, ou seja,  $v_0 = v_k$ , uma trilha é definida como uma caminhada em que todas as arestas são distintas e um caminho é uma trilha em que todos os vértices são distintos. E um círculo será uma trilha fechada ao quais todos os vértices são distintos, menos os vértices  $v_0$  e  $v_k$ .*

Usando a noção de caminho, é possível definir um grafo como conectado quando houver um caminho entre cada par de seus vértices. Formalmente:

**Definição 2.8.** *Dois vértices distintos  $u$  e  $v$  em um grafo  $G$ , estão conectados se houver um caminho  $c(u, v)$  em  $G$ . De modo que  $G$  será conectado se todos os pares de vértices distintos também estiverem conectados.*

Intuitivamente isso significa que um grafo também pode consistir de uma coleção de componentes, vértices e arestas, de modo que cada componente seja um subgrafo conectado. De forma precisa, segue-se:

**Definição 3.9.** *Um subgrafo  $H$  de  $G$  será denominado um componente de  $G$  se  $H$  for conectado de modo que seus elementos não estiverem conectados a outro subgrafo de  $G$ . O número de componentes de  $G$  é denotado  $\omega(G)$ .*

O conceito de conectividade em um grafo é notável quando é necessário o estudo de robustez de redes.

### 2.1.8 Grafos direcionados ou dígrafos

Até agora tratamos de grafos com vértices conectados por arestas especificadas pelo par não ordenado de vértices  $(u, v)$  para o caso de grafos simples. Quando temos direcionalidade em arestas, tratamos de grafos direcionados que possuem propriedades especiais. Para tanto, temos:

**Definição 2.10.** *Um grafo direcionado, ou dígrafo,  $D$  consiste numa coleção de vértices  $V$  e uma coleção de arcos  $R$  de modo que denotamos  $D = (V, R)$ . De modo que cada arco  $r = (\overline{u, v})$  conecta um vértice  $u$  ao vértice  $v$ . O vértice  $u$  é denominado origem de  $r$  e  $v$  sua extremidade.*

O grafo subjacente  $G(D)$  do dígrafo  $D$  é obtido pela troca do arco  $r = (\overline{u, v})$  por uma aresta não direcionada. De modo que se um grafo possui  $m$  arestas, haverá  $2^m$  orientações possíveis das arestas do dígrafo. Analogamente a grafos simples, os vizinhos de um vértice em dígrafos têm significado importante. Para tanto consideremos:

**Definição 2.11.** *Considere um dígrafo  $D$  e um de seus vértices  $v \in V(D)$ . O conjunto de vizinhos de entrada  $N_{in}(v)$  de  $v$  consiste no número de vértices adjacentes que têm extremidade coincidente com  $v$ . Analogamente, o conjunto de vizinhos de saída  $N_{out}(v)$  consiste no número de vizinhos de  $v$  em que as origens de seus arcos coincidem. Formalmente:*

$$N_{in}(v) \stackrel{\text{def}}{=} \{w \in V(D) | v \neq w, \exists r = (\overline{w, v}) : r \in R(D)\}, \quad (2.9)$$

$$N_{out}(v) \stackrel{\text{def}}{=} \{w \in V(D) | v \neq w, \exists r = (\overline{v, w}) : r \in R(D)\}. \quad (2.10)$$

Interessantemente, o número de vizinhos do vértice  $v$ ,  $N(v)$ , será a união entre os vizinhos de saída e de entrada

$$N(v) \stackrel{\text{def}}{=} N_{in}(v) \cup N_{out}(v). \quad (2.11)$$

Adicionalmente, um dígrafo é denominado estrito se não possuir alças e nenhum par de arcos é incidente aos mesmos dois vértices. A noção de dígrafo estrito é análoga ao



conceito de grafo simples. E de forma também análoga, é possível definir o grau de um vértice para um dígrafo:

**Definição 2.12.** Para um vértice  $v$  de um dígrafo  $D$ , o número de arcos com extremidades coincidentes em  $v$  é denominado grau de entrada  $\delta_{in}(v)$  de  $v$ . Analogamente, o grau de saída  $\delta_{out}(v)$  de  $v$  é o número de arcos em que as origens coincidem com  $v$ .

Desta definição, segue-se com uma conclusão inevitável:

**Teorema 2.2.** Para qualquer dígrafo  $D$ , a soma dos graus de entrada assim como a soma dos graus de saída será igual ao número de arcos no dígrafo:

$$\sum_{v \in V(D)} \delta_{in}(v) = \sum_{v \in V(D)} \delta_{out}(v) \equiv |R(D)|. \quad (2.12)$$

A representação do dígrafo por meio de uma matriz também é possível, caracterizando a matriz de adjacência de um dígrafo,  $A$ , em que  $A[i, j]$  é equivalente ao número de arcos que une um vértice qualquer  $v_i$  à  $v_j$ . Em contraste com o conceito de matriz de adjacência de um grafo simples, é possível identificar algumas propriedades específicas da matriz de adjacência de um dígrafo:

1. Um dígrafo  $D$  será estrito se, e somente se, para todo  $i$  e  $j$  sejam satisfeitas as condições:  $A[i, j] \leq 1$  e  $A[i, i] = 0$ . Em outras palavras, pode haver no máximo um arco que une um vértice qualquer  $v_i$  a  $v_j$  e não pode haver nenhum arco em que sua extremidade e origem coincidam no mesmo vértice;
2. Para cada vértice  $v_i$ , há um  $\sum A[i, j] = \delta_{out}(v_i)$  e um  $\sum A[j, i] = \delta_{in}(v_i)$ . Em outras palavras, a soma das entradas na linha  $i$  da matriz, corresponde ao grau de saída do vértice  $v_i$ ; onde a soma das entradas da coluna  $i$  corresponde ao grau de entrada do vértice  $v_i$ .

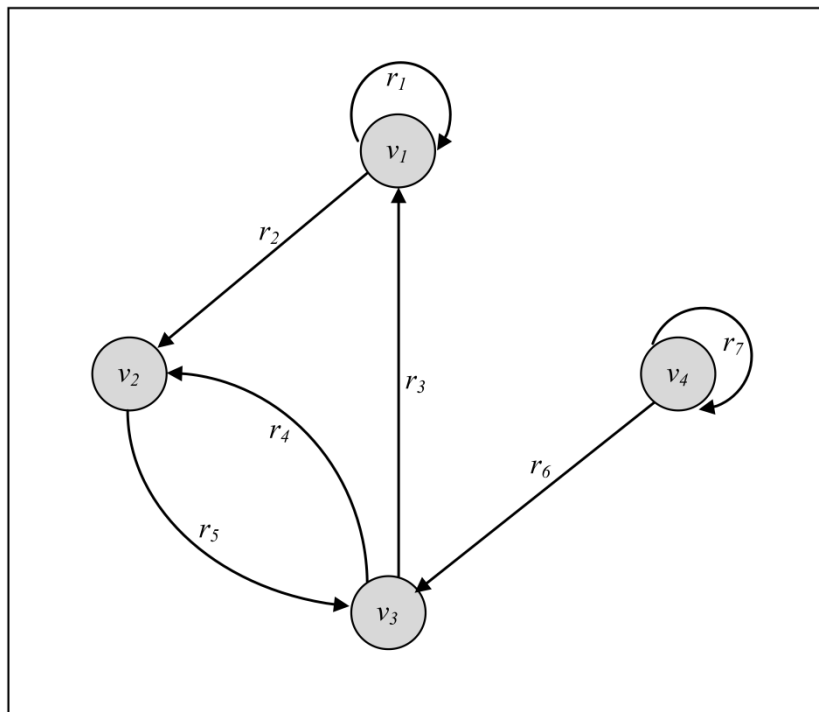
É de importância notar que as matrizes de adjacência de dígrafos, nem sempre são simétrica como o são em grafos simples; ou seja,  $A[i, j] \neq A[j, i]$ . A seguir, um exemplo de um dígrafo e sua matriz de adjacência associada:

Tabela 2.3 Matriz de adjacência digrafo da Figura 2.2.

|          | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $\Sigma$ |
|----------|-------|-------|-------|-------|----------|
| $v_1$    | 1     | 1     | 0     | 0     | 2        |
| $v_2$    | 0     | 0     | 1     | 0     | 1        |
| $v_3$    | 1     | 0     | 0     | 0     | 1        |
| $v_4$    | 0     | 0     | 1     | 1     | 2        |
| $\Sigma$ | 2     | 1     | 2     | 1     | 6        |

Fonte: autor.

Figura 2.2 Exemplo de um dígrafo.



Fonte: autor

Similarmente, é possível representar um dígrafo por meio de sua matriz de incidência. Mas neste caso,  $\mathbf{M}[i, j]$  representa a matriz que denota a incidência de algum vértice  $v_i$  ao arco  $r_j$ . Os valores das entradas desta matriz devem seguir as condições:

$$\mathbf{M}[i, j] = \begin{cases} 1, & \text{se o vértice } v_i \text{ for incidente à origem de } r_j. \\ -1, & \text{se o vértice } v_i \text{ for incidente à extremidade de } r_j. \\ 0, & \text{se não houver nenhuma incidência.} \end{cases}$$

Como um dígrafo pode possuir alças, ou seja, arcos que possuem incidências num mesmo vértice, a representação do dígrafo por meio da matriz de incidência é pouco comum. Outro conceito importante em dígrafos é o de conectividade. Para definir conectividade em dígrafos, é preciso uma definição equivalente a caminho em grafos:

**Definição 2.13.** *Considere um dígrafo  $D$ . Uma caminhada direcionada  $(v_0, v_k)$  em  $D$  é a sequência alternante  $[v_0, r_0, v_1, r_1 \dots v_{k-1}, r_{k-1}, v_k]$  de vértices e arcos de  $D$  com  $r = \langle \overrightarrow{v_i, v_{i+1}} \rangle$ . Uma trilha direcionada é a caminhada direcionada em que todos os arcos da sequência são distintos; e um caminho direcionado é a trilha direcionada em que todos os vértices são distintos. Com isso, um círculo direcionado é a trilha direcionada em que todos os vértices são distintos com exceção dos vértices  $v_0$  e  $v_k$ .*

É notável que os conceitos de caminhada, trilhas, caminhos e círculos são completamente análogos para o caso de grafos simples. Com isso é possível definir o conceito de conectividade em dígrafo:

**Definição 2.14.** *Um dígrafo será fortemente conectado se houver um caminho direcionado entre cada par de vértices distintos de  $D$ . Analogamente, um dígrafo será fracamente conectado se seu grafo subjacente for conectado.*

O conceito de conectividade em dígrafos é de grande importância no estudo de fluxo de informação em redes complexas como será tratado em um dos capítulos subsequentes acerca dos Processos Markovianos.

## 2.2 Redes complexas: análise e caracterização

De modo grosseiro, redes complexas são sinônimas de grafos, mas há uma diferença subjetiva que é relevante: diferentemente de grafos, redes complexas são construídas por dados coletados na natureza e sua composição estrutural tende a ser não trivial e carregada de significado. Dada estas características, modelos baseados em redes complexas permitem a descrição de diversos fenômenos na natureza e na sociedade. Alguns exemplos podem ser citados, como: a constituição celular e suas reações metabólicas, a Internet, a rede de

roteadores e computadores fisicamente conectados, a “World Wide Web” com sua rede de “hyperlinks”, espécies que compõem redes tróficas, redes de interação inseto-planta e assim por diante. Inicialmente, estes estudos famosos foram descritos como redes aleatórias e então associadas a padrões não triviais como os fenômenos do “Small-World” e “Free Scale”, onde a constituição estrutural da rede é significativa.

Nesta seção serão introduzidas as principais redes estudadas na literatura assim como as ferramentas de análise da composição estrutural de redes complexas mais utilizadas.

### **2.2.1 Considerações gerais sobre redes complexas**

Associar sistemas físicos a redes complexas permite uma variedade de novas análises tanto para finalidades tecnológicas quanto para a ciência básica. O modelo em redes complexas dá margem para a análise de sua constituição estrutural, frequentemente denominada topologia (não no sentido da álgebra moderna) que esta associada com a complexidade das relações entre vértices e arestas. Um exemplo de aplicação deste tipo de modelo: consideremos espécies biológicas, e a relação entre elas relações tróficas. Se todos os elementos estiverem corretamente associados, é possível, por meio da teoria de grafos, identificar quais destas espécies são mais importantes para manter a integridade total da rede complexa, isto é, da cadeia trófica. Em última análise, a modelagem de fenômenos naturais usando redes complexas, permite novos tipos de análises de modo a gerar conclusões que o reducionismo, por si só, é insuficiente.

Até então, a física tem sido o maior beneficiário da proposta reducionista desenvolvendo um grande conjunto de poderosas ferramentas de descrição e previsão de comportamento de sistemas físicos mais simples. É graças a esta posição filosófica (*i.e.*, reducionismo) que conhecemos um pouco mais, por exemplo, dos fenômenos do magnetismo em materiais com esta propriedade; o magnetismo, segundo essa visão, é um comportamento que depende do comportamento individual de cada spin de todos os elementos moleculares da rede atômica do material. É possível dizer, para este caso, que a soma dos comportamentos individuais determinam o comportamento do todo. Outros exemplos de comportamento coletivo, na Física, são os Condensados de Bose-Einstein [28] ou mesmo a superfluidez, para citar exemplos quânticos que refletem padrões macroscópicos.

O sucesso de modelos reducionistas, isto é, de modelos que descrevem o comportamento global pela soma direta dos comportamentos individuais, é baseado intrinsecamente na natureza do sistema modelado. Ou seja, a forma com que os constituintes

deste sistema interagem entre si denotando simplicidade: esta é a premissa para o sucesso reducionista. Exemplos de sistemas simples são: um oscilador, um pêndulo, uma roda em rotação ou um planetóide orbitando um ponto. Estes exemplos suscitam que não há ambiguidade nas relações de grandezas que determinam seu comportamento. Analogamente, há pouca interação cruzada entre elementos, e quando há, como é o caso do campo elétrico e/ou magnético, a interação é determinada unicamente pela distância física excetuando-se outras grandezas, como carga elétrica e densidade de corrente. Nestas condições, o método reducionista é ideal para a previsão e descrição de fenômenos físicos.

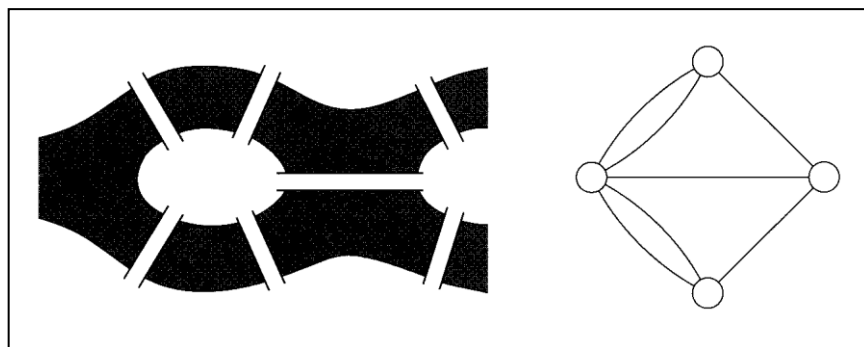
Por outro lado, há uma quebra no domínio reducionista quando os comportamentos de alguns sistemas levam em conta relações não triviais, por exemplo, independentes da distância física entre suas partes. Alguns sistemas possuem ambiguidade acerca de quais grandezas realmente determinam seu comportamento global e em que tipo de dependência temporal, isto é, dinâmica. A todos estes casos dá-se o nome coletivamente de sistemas complexos, aos quais idealmente os modelamos como redes complexas denotadas matematicamente como grafos. É conveniente usar grafos para representar sistemas reais com interações complexas, pois no momento de modelar a não trivialidade interacional já se encontra implícita no modelo. Nota-se concomitantemente que, nos últimos anos, houve um aumento na criação e uso de ferramentas da mecânica estatística associadas a sistemas complexos (*i.e.*, grafos) devido ao atual interesse das ciências naturais na interdisciplinaridade.

### **2.2.2 Grafos e redes complexas: histórico e desenvolvimento**

Tradicionalmente, o estudo de redes complexas tem sido domínio, em verdade, da teoria de grafos [11]. Nos últimos anos, diversas classes de cientistas como físicos, químicos, biólogos, matemáticos e sociólogos, têm construído uma teoria de redes complexas baseando-se em redes reais e resultados analíticos em teoria de grafos [12]. O estudo de redes tem um histórico de longa data influenciada pela matemática e ciências da época. Em 1736, um renomado matemático, Leonard Euler, se interessou pelo enigma associado ao problema de ponte de Königsberg. Nesta cidade, sete pontes conectavam as massas de terra circundadas por rios na região da antiga Prússia, e havia um ditado comum: “Há algum caminho único no qual seja possível atravessar as sete pontes apenas uma vez?” Muitos tentaram responder essa pergunta até que Euler provar que não há meios para atravessar as sete pontes passando por elas uma única vez. A prova que pode parecer trivial nos tempos atuais, não era tão óbvia em

1736, pois a resposta depende do conceito de grafo até então inexistente. Grafo, o objeto matemático composto por pontos e linhas que conectam pontos. Apesar de ser um conceito simples, permite poderosas formas de resolução de problemas como o caso da ponte de Königsberg. Porém, dado o conceito de grafo, o problema pode ser reformulado neste novo aspecto matemático do problema: a questão é se há um caminho euleriano na rede associada às pontes. Um caminho euleriano, em teoria de grafos, é precisamente a sequência de vértices que passa por cada aresta apenas uma vez no grafo todo. Euler observou no grafo associado ao problema das pontes (figura 2.3), que não há caminho euleriano [9]. Isso significa que pode haver no máximo dois vértices com número ímpar de arestas conectadas. Como no grafo associado às pontes, todos os vértices têm número ímpar de arestas incidentes, então o problema não tem solução. Esta interpretação só foi possível com o uso do conceito de grafo e de caminho euleriano e caracterizou o marco inicial desta nova ramificação da matemática: a teoria de grafos. Desde então a teoria de grafos têm sido a principal teoria que estuda redes reais. Os elementos do grafo vértice e arestas, pareados, podem assumir qualquer configuração, como: pessoas e relações humanas, computadores e redes físicas, compostos químicos e reações químicas, espécies biológicas e relações tróficas, etc.

Figura 2.3 Esquema das Pontes de Königsberg e seu grafo associado: os vértices são porções de terra e as pontes as arestas.



Fonte: figura adaptada de Newman et al. 2006, referência [33].

O método analítico permeado pela teoria de grafos é poderoso, sendo a possibilidade de aplicação em diversas áreas do conhecimento a prova desta afirmação. Em detrimento da diversidade de aplicação dos métodos analíticos gerados pela teoria de grafos e então teoria de redes complexas, em 1950, houve um florescimento no interesse de geradores de quantificadores especialmente nas áreas de ciências sociais e antropologia. Onde os padrões de relações entre pessoas puderam ganhar uma profundidade analítica maior [13].

Posteriormente, estes tipos de estudos se espalharam para outras áreas de pesquisa em que o conceito de grafo pudesse ser aplicado.

Os grafos randômicos foram primeiramente estudados pelos matemáticos húngaros Paul Erdős e Alfréd Rényi, em 1959; de acordo com seu modelo, considera-se um número  $N$  de vértices iniciais de modo a conectar cada par de vértices com uma probabilidade  $p$ , criando, no processo, um grafo de aproximadamente  $pN(N - 1)/2$  arestas distribuídas randomicamente [14]. Este foi o primeiro modelo proposto como descritor para sistemas complexos reais, porém era sabido que a não trivialidade das interações entre os vértices destes grafos não fazia sentido na condição complexa dos sistemas reais. Era sabido que havia relações muito complexas entre as componentes de uma célula ou de um ecossistema, citando exemplos: relações tróficas complexas, relações de coevolução, relações inseto-planta, etc. Numa instância primária, é apreensível que os sistemas reais devem ter um princípio organizacional que reflete no modelo do grafo que o descreve. Este elemento é totalmente perdido no modelo de grafos randômicos à medida que as ligações têm mesma probabilidade de ocorrer em quaisquer vértices, imagina-se que poucos fenômenos naturais se adaptem a esta fenomenologia estrutural.

Apesar da impossibilidade de se modelar honestamente um sistema real por meio do modelo de Erdős e Rényi, esta aproximação ainda se torna útil se é requerido quantificar o grau de complexidade de uma rede construída empiricamente por coleta de dados direta em detrimento da mesma rede (*i.e.*, mesmo número de vértices) criada randomicamente. De modo que com os grafos, calcula-se a diferença entre suas distribuições de graus possibilitando muitos quantificadores de complexidade [12].

Interessantemente, a ciência computacional e suas tecnologias associadas tiveram papel fundamental na construção do corpo teórico que hoje é denominada Teoria de Redes Complexas. A primeira importante contribuição foi a possibilidade de armazenamento de grandes quantidade de informação. Em segundo lugar foi a possibilidade de analisar vários dados ao mesmo tempo. E finalmente, em terceiro lugar, a quebra do paradigma reducionista em prol da ciência multidisciplinar. Com estas considerações, as formas de análise propostas pela teoria de grafos e então teoria de redes complexas, são muito necessárias quando é pretendido o estudo da complexidade em fenômenos naturais e humanos.

Desta preocupação com a relação do modelo com a realidade, três conceitos fundamentais foram proeminentes no atual estado da arte da teoria de redes complexas. A primeira delas é o fenômeno “small world”, que enfatiza o fato de que apesar de as redes poderem ser grandes, os caminhos entre quaisquer dois vértices tendem a ser pequenos. A

segunda é a tendência de redes reais estarem constituídas por “clusters”, ou aglomerados. Isso é caracterizado pela tendência de os vértices se organizarem em grupos de vértices com alto grau entre eles, enquanto que as relações entre estes grupos são menos densas, ou seja, relação de baixo grau. A terceira é a forma da distribuição de grau: foi verificado extensamente que a maior parte das redes reais tem distribuição de grau seguindo uma lei de potência, o que caracteriza o fenômeno conhecido como “free scale”. Este fenômeno, em verdade, pode ser entendido como a tendência de a rede possuir poucos vértices muito conectados, isto é, alto grau, e muitos vértices pouco conectados. A seguir são tratados os modelos de redes complexas mais detalhadamente.

### 2.2.3 Teoria de redes randômicas

Como comentado anteriormente, classicamente o modelo de grafos foi estudado por Erdos e Rényi [14-16]. Estes matemáticos definiram o grafo aleatório como um grafo com  $N$  vértices rotulados e conectados por  $n$  arestas as quais são escolhidas randomicamente sob a probabilidade  $pN(N-1)/2$ , [14]. De modo que as configurações possíveis para um grafo desse tipo será equivalente à combinação  $C_{[N(N-1)/2]}^n$ , e cada configuração deve ser equiprovável. Uma definição alternativa de grafo randômico é o modelo binomial. Neste caso, consideramos inicialmente  $N$  vértices, de modo que cada par de vértices pode ser conectado com uma probabilidade  $p$ . Isso implica que o número total de arestas é uma variável aleatória de modo que o valor esperado seja  $A(N) = p[N(N-1)/2]$ . E se for considerado um grafo inicial  $G_0$  composto por  $N$  vértices e  $n$  arestas, a probabilidade de obtê-lo no processo de criação de grafos será  $P(G_0) = p^n(1-p)^{N(N-1)/(2-n)}$ . A figura 2.4 ilustra um exemplo de grafos criados por meio do modelo de Erdos e Rényi com probabilidades  $p$  crescentes.

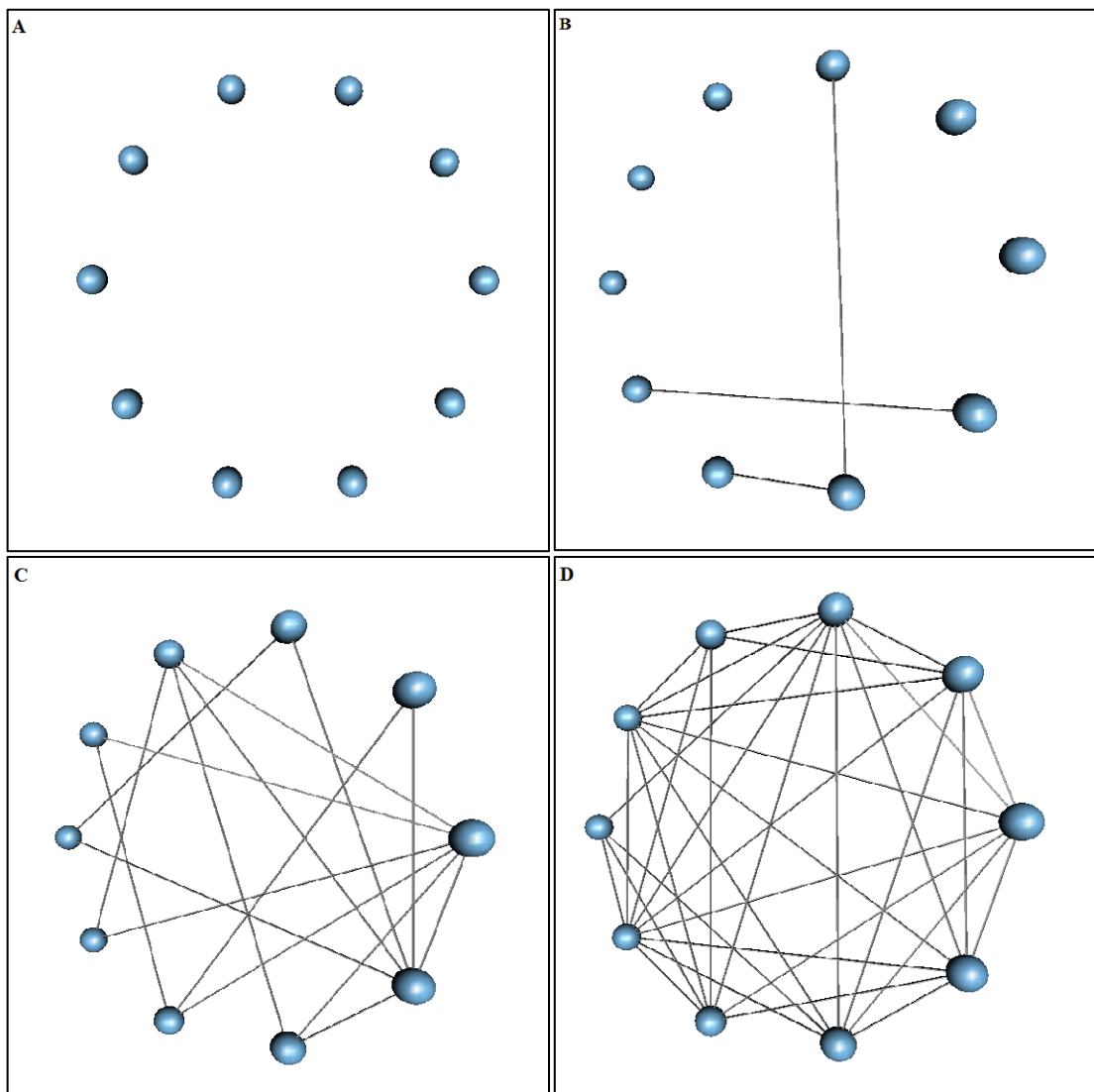
A teoria de grafos randômicos estuda, em última análise, as propriedades do espaço de probabilidades associados a grafos com  $N$  vértices à medida que  $N \rightarrow \infty$ . A construção de grafos por meio da adição de novas arestas é denominada evolução do grafo, como demonstrado na figura 3.4: iniciou-se o grafo com  $N$  vértices isolados de modo que o grafo se organiza com o acréscimo crescente de arestas que é proporcional ao aumento de  $p$ . O grafo em evolução corresponderá a diversos estágios, eventualmente alcançando  $p = 1$ , onde o número máximo de arestas é alcançado e o grafo se torna completo.

Um dos principais objetivos desta teoria é determinar a probabilidade  $p$  em que uma propriedade especial do grafo deve emergir. Com isso, a grande colaboração de Erdős e Rényi



neste âmbito, foi a descoberta de muitas propriedades importantes na evolução de grafos aleatórios, como o surgimento da componente gigante e os subgrafos [16]. Em termos mais formais, isso é o mesmo que dizer que se há uma propriedade  $Q$  do grafo, o mesmo emergirá com certa probabilidade  $p$ , de modo que a maior parte dos grafos possa exprimir essa afirmação.

Figura 2.4 Exemplo de processo de criação de grafos por meio do modelo de Erdos-Rényi. A)  $N = 10$ ,  $p = 0$ ; B)  $N = 10$ ,  $p = 0,1$ ; C)  $N = 10$ ,  $p = 0,3$ ; D)  $N = 10$ ,  $p = 0,7$ .



Fonte: autor

De modo que haja, conseqüentemente, uma probabilidade crítica  $p_c(N)$  de surgimento de uma propriedade à medida que a rede evolui. Se  $p(N)$  crescer em menor escala que  $p_c(N)$  à medida que  $N \rightarrow \infty$ , então quase todo grafo com probabilidade de emergência  $p(N)$  de novas

propriedades falha em gerar  $Q$ . Adicionalmente, a probabilidade de um grafo com  $N$  vértices e probabilidade conexão  $p = p(N)$  de ter a probabilidade  $Q$  deve satisfazer o seguinte limite:

$$\lim_{N \rightarrow \infty} P_{N,p}(Q) = \begin{cases} 0 & \text{se } \frac{p(N)}{p_c(N)} \rightarrow 0 \\ 1 & \text{se } \frac{p(N)}{p_c(N)} \rightarrow \infty \end{cases}. \quad (2.13)$$

É de importância evidenciar que em redes complexas, o número de possíveis combinações de vértices aumenta exponencialmente com o aumento do número de vértices. Conseqüentemente, na teoria de grafos randômicos a probabilidade de ocupação é definida como a função que depende do tamanho do sistema. Neste âmbito,  $p$  representa a fração de arestas que estão presentes no rol de possibilidades  $N(N - 1)/2$ . Grafos grandes de mesmo  $p$  irão conter mais arestas, e conseqüentemente a propriedade como círculos irão ocorrer com mais frequência do que em grafos menores de mesma probabilidade. Isso significa que para muitas propriedades  $Q$  em grafos randômicos não há um único limiar- $n$  independente, mas uma função limite definida que depende do tamanho do sistema e da probabilidade  $p_c(N \rightarrow \infty) \rightarrow 0$ . Porém é notável que o grau médio  $\langle k \rangle$  do grafo seja dado pelo desenvolvimento

$$\langle k \rangle = \frac{2n}{N},$$

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

$$\langle k \rangle \sim pN. \quad (2.14)$$

O grau médio dado por 2.14 tem um valor crítico que é independente do tamanho do sistema. Na próxima seção o processo de emergência de propriedades estruturais em redes randômicas ficará mais claro.

#### 2.2.4 Subgrafos em redes randômicas

A primeira propriedade  $Q$  que emerge, estudada por Erdős e Rényi em suas redes randômicas, são os subgrafos. Se considerarmos um grafo qualquer  $G_0$  composto por um

conjunto de  $V_0$  vértices e um conjunto de  $A_0$  arestas; de modo que  $G_0$  seja um subgrafo de  $G = \{V, A\}$ , então todo  $V_0 \in V$  e  $A_0 \in A$ . Exemplos simples de subgrafos são constituições como ciclos, árvores, triângulos, quadrados, etc. É possível dividir os subgrafos simples em fechados e abertos: os fechados são normalmente denominados como ciclos e são um “loop” de  $k$  arestas de modo que a cada duas arestas consecutivas haja apenas um vértice comum. Deste conceito derivamos os triângulos que podem ser concebidos como ciclos de ordem 3, quadriláteros que são ciclos de ordem 4, pentágonos de ordem 5 e assim por diante. Analogamente, têm-se os subgrafos abertos que são as árvores que não formam “loops”: mais precisamente uma árvore, digamos, de ordem  $k$  é a árvore que tem  $k$  vértices e  $k - 1$  arestas e nenhuma de suas extensões forma ciclos.

A probabilidade, assim como o número de estruturas complexas, é possível de se calcular [17]: considerando um grafo randômico de modo que  $G = G_{N,p}$ . De modo que consideremos um segundo grafo  $F$  menor, que consiste de  $k$  vértices e  $l$  arestas. Em princípio, o grafo randômico  $G$  pode conter, por probabilidade, grafos isomórficos a  $F$ . Para tanto, podemos estimar quantos destes grafos podem existir em  $G$ . Os  $k$  vértices de  $F$  podem ser escolhidos do número total de possíveis vértices  $N$  na proporção da combinação  $C_N^k$ ; e as arestas de sua constituição dependem da probabilidade com que  $G$  evoluiu sob  $p$ , então as arestas de  $F$  são formadas na proporção de  $p^l$ . Ainda assim, é possível permutar os elementos  $k$  escolhidos e potencialmente obter  $k!$  novos grafos de constituição  $F$ . Mais rigorosamente, são  $k!/a$  configurações, onde  $a$  é o número de grafos isomórficos entre os possíveis grafos. Com isso, o número esperado de subgrafos  $F$  contidos em  $G$  pode ser escrito, no limite de  $N$  tendendo ao infinito, como

$$E(X) = C_N^k \frac{k!}{a} p^l \simeq \frac{N^k p^l}{a}. \quad (2.15)$$

A relação acima sugere que o número de subgrafos  $X$  é aproximadamente igual a  $E(X)$  de modo a indicar que quando  $p(N) = N^{k/l} \rightarrow 0$  à medida que  $N \rightarrow 0$ ,  $E(X) \rightarrow 0$ . Porém se  $p(N) = cN^{k/l}$ , então o número médio de subgrafos é um número finito denotado por  $\lambda = c^l/a$  que indica a probabilidade crítica citada anteriormente. Interessantemente, algumas relações de probabilidade para este caso são possíveis identificadores de estruturas emergentes num grafo [12, 17]:

- i. A probabilidade de  $G$  conter uma árvore de ordem  $k$  será

$$p(N) = cN^{-\frac{k}{(k-1)}}. \quad (2.16)$$

- ii. A probabilidade de  $G$  conter um ciclo de ordem  $k$  será

$$p(N) = cN^{-1}. \quad (2.17)$$

- iii. E a probabilidade de  $G$  conter um subgrafo completo (todos os vértices são vizinhos uns dos outros) será

$$p(N) = cN^{-\frac{2}{(k-1)}}. \quad (2.18)$$

Estes resultados são importantes, pois quando evoluímos um grafo randômico segundo os parâmetros  $N$  e  $p$ , é possível estimar a probabilidade de ocorrência de estruturas emergentes, permitindo suas previsões de emergência. Algumas estruturas surgem com mais frequência dada certa dependência em  $N$  [16, 17]. A discussão no próximo tópico esclarece este fenômeno.

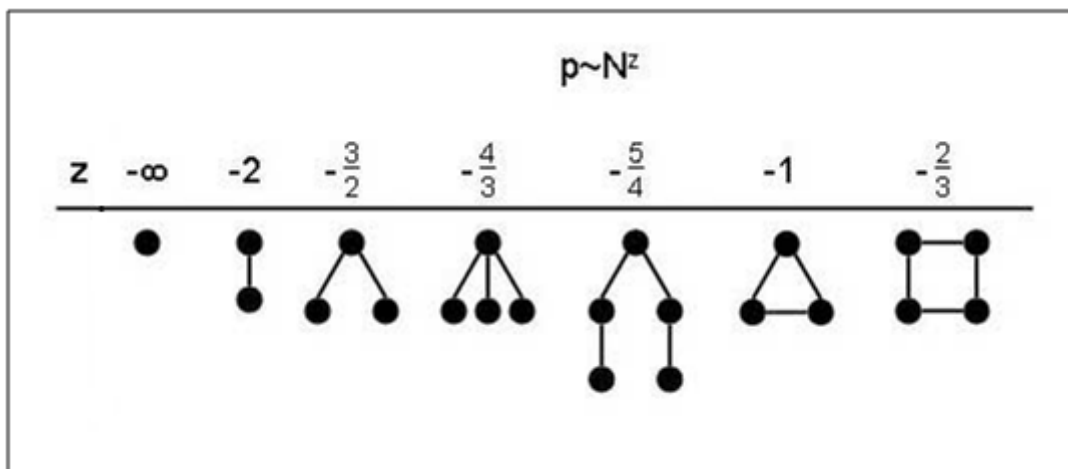
### 2.2.5 Evolução de redes randômicas

É possível encontrar os resultados dados no tópico anterior por meio da aproximação de evolução de grafos randômicos. Para tanto, consideremos um grafo com  $N$  vértices de modo a assumir uma probabilidade de conexão de  $p(N)$  que é proporcional à um certo  $N^z$ , onde  $z$  é um parâmetro de ajuste que toma qualquer valor entre  $-\infty$  e 0. Para  $z$  menor que  $-3/2$ , quase todos os grafos contêm apenas vértices e arestas isoladas. Quando o valor de  $z$  ultrapassa  $-3/2$ , árvores de ordem 3 emergem. Quando  $z$  alcança o valor de  $-4/3$ , árvores de ordem 4 emerge. De modo que a medida que  $z$  alcança o valor de  $-1$ , o grafo irá conter árvores de ordens maiores [17]. Porém, a medida que  $z > -1$ , de modo que o grau médio do grafo  $\langle k \rangle = pN \rightarrow 0$  e  $N \rightarrow \infty$ , o grafo são uniões disjuntas de árvores e ciclos estão ausentes.

Paradoxalmente, quando  $z$  passa o valor de  $-1$ , a probabilidade assintótica de ciclos de todas as ordens passa de 0 para 1. Ciclos de ordem 3, podem ser interpretados como

subgrafos completos de ordem 3. Analogamente, subgrafos completos de ordem 4 emergem em  $z = -2/3$ , e a medida que  $z$  diminui, mais ciclos de maiores ordens tendem a emergir. Para  $z \rightarrow 0$ , o grafo irá conter todos os subgrafos completos de ordens finitas. A figura 2.5 expressa pictoricamente estas possibilidades.

Figura 2.5 As probabilidades limiaries em que diversos subgrafos emergem num processo de evolução de um grafo randômico, em que  $p$  é a probabilidade de conexão entre vértices,  $N$  é o número de vértices no grafo e  $z$  é um parâmetro de sintonia que varia entre  $-\infty$  e 0.



Fonte: figura adaptada de Barabási et al. (2002) referência [32].

### 2.2.6 Teoria de grafos “small world”

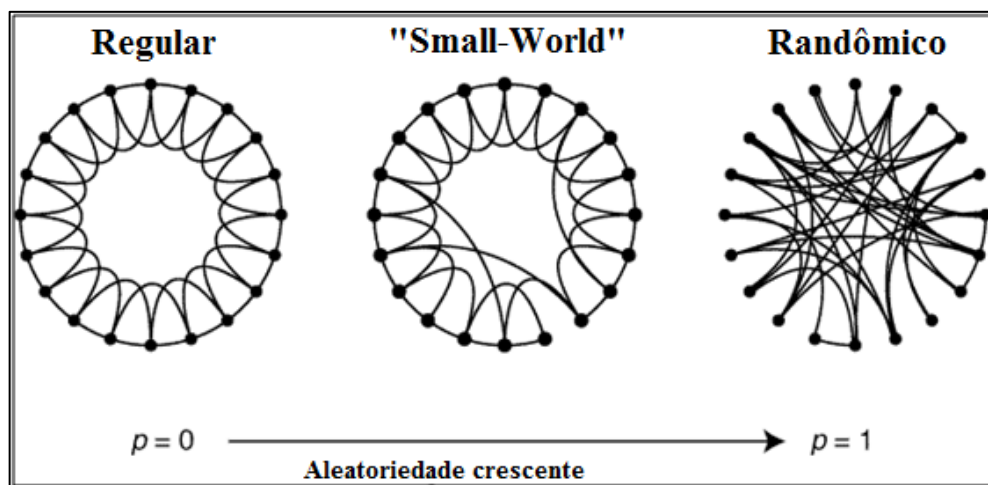
Modelo proposto inicialmente por Watts e Strogatz [18], considera a construção da rede num modelo que depende apenas de um parâmetro que interpola entre um estado inicial ordenado, ciclos de diversas ordens, e grafos randômicos. O procedimento geral pode ser sintetizado no seguinte esquema:

- i. Começando com um grafo do tipo anel com  $N$  vértices conectados com seus primeiros  $k$  vizinhos, ou seja,  $k/2$  vizinhos. Para manter a homogeneidade, mas ao mesmo tempo certa conectividade no grafo, a seguinte condição deve ser satisfeita

$$N \gg K \gg \ln(N) \gg 1;$$

- ii. Em seguida, reconecta-se randomicamente cada aresta, do grafo anel dado acima, com uma probabilidade  $p$  de modo que a autoconexão e a duplicação de arestas estejam excluídas. Este passo introduz atalhos na proporção de  $pNK/2$ , de modo a conectar vértices que não são os vizinhos mais próximos de grau  $k$ . Variando  $p$ , é possível controlar a tração entre estados de ordem  $p = 0$  e aleatoriedade na rede  $p = 1$ .

Figura 2.6 O modelo de Watts-Strogatz; mostrando a evolução da rede com o aumento no número atalhos de rede proporcional à probabilidade de reconecção  $p$ .



Fonte: figura adaptada de Watts et al. (1998) referência [18].

A Figura 2.6 exprime os passos desse procedimento desenvolvido por Watts e Strogatz, também denominado modelo de Watts-Strogatz. Este modelo tem suas origens nos estudos de sistemas sociais nos quais a maior parte das pessoas são amigos de seus vizinhos imediatos. Aqui o termo vizinho está associado à proximidade física, como: vizinhos da mesma rua, colegas de trabalho, amigos de amigos, e assim por diante. Porém, todas as pessoas têm um ou dois amigos distantes, ou seja, pessoas de outros países, conhecidos antigos, etc. Estes são a representação dos atalhos na rede social dada pelo modelo de Watts-Strogatz [12, 18]. Para entender a coexistência dos atalhos na rede e do processo de agrupamento, se faz necessário entender o comportamento do coeficiente de aglomeração  $C(p)$  e a distância de caminho médio  $\ell_G$  como função da probabilidade de reconecção  $p$ . A distância de caminho médio para um grafo genérico pode ser calculada como

$$\ell_G = \frac{1}{n(n-1)} \sum_{i,j} d(v_i, v_j). \quad (2.19)$$

onde  $d(v_i, v_j)$  é a menor distância entre os vértices  $v_i$  e  $v_j$ , e  $n$  o número de vértices no grafo. Analogamente, se considerarmos que um vértice  $i$  qualquer possua  $K_i$  vizinhos, verificamos combinatoriamente que o número total de vizinhos será  $K_i(K_i - 1)/2$ . Seja o número real de arestas do vértice  $i$ , com seus vizinhos dado por  $N_i$ , o coeficiente de aglomeração pode ser escrito como

$$C_i = \frac{2N_i}{K_i(K_i - 1)}. \quad (2.20)$$

O coeficiente de aglomeração da rede toda é simplesmente a média dos coeficientes individuais dos  $N$  vértices da rede. Para um grafo anel temos que  $\ell(0) \simeq N/2K \gg 1$  e  $C(0) \simeq 3/4$ . Adicionalmente  $\ell$  é linearmente dependente com o tamanho da rede e o coeficiente de aglomeração tende a ser grande. Por outro lado, se  $p \rightarrow 1$  o modelo converge para um grafo randômico no qual  $\ell(1) \sim \ln(N)/\ln(K)$  e  $C(1) \sim K/N$ . Neste caso  $\ell$  depende logaritmicamente de  $N$  e o coeficiente de aglomeração diminui com  $N$ . Estes casos limites sugerem que altos valores de  $C$  estão sempre associados a altos valores de  $\ell$ , e o inverso também é verdadeiro [18].

### 2.2.7 Teoria de redes livre de escala

É defensável que muitas das redes reais estudadas atualmente, com um número de vértices considerável, têm um caráter especial acerca de sua distribuição de graus: ela segue uma lei de potência para altos valores de  $k$  (*i.e.*, grau do vértice) e uma lei exponencial para a cauda da distribuição [19]. Os modelos que geram grafos do tipo randômico ou “small-world” não reproduzem essa característica [12]. A distribuição de graus é dada pelas probabilidades  $p(k)$ , que representam a probabilidade de um dado vértice ter grau  $k$ , então a distribuição do tipo lei de potência deve satisfazer

$$p(k) \sim k^{-\gamma}. \quad (2.21)$$

para  $\gamma$  constante e positivo. Ser livre de escala, em termos grosseiros, significa que há poucos nós muito conectados e muitos nós pouco conectados; para tanto, estes autores assumem características fundamentais que diferem muito do modelo de Erdős-Rényi e de Watts-Strogatz. Dada esta nova propriedade, segue a questão: qual o mecanismo responsável pela emergência das redes livre de escala? Este modelo, conhecido como modelo de Barabási-Albert, sugere que a forma com que a rede evolui deve determinar a característica livre de escala.

Barabási e Albert, em seu modelo de rede livre de escala [19], argumentam que a natureza deste fenômeno está enraizada em dois mecanismos genéricos presentes em muitas redes reais. O primeiro deles é o fato de que a maior parte das redes cresce com o tempo, ou seja,  $N$  é função de  $T$ . Por exemplo, o número de adição de novas “web pages” no WWW aumenta exponencialmente no tempo [12]. O segundo fator é o caso da dependência na probabilidade de existir uma aresta entre dois vértices quaisquer. Barabási e Albert defendem que em redes reais, há uma tendência preferencial de conexão entre vértices de modo a refletir a não trivialidade na rede como um todo. Estas duas características não são assumidas pelos modelos de Erdos-Renyi e Watts-Strogatz, pois  $N$  é constante com o tempo e a probabilidade de conexão entre dois vértices distintos são funções de fatores aleatórios ou arbitrários.

Estes dois aspectos apresentados no modelo de Barabási-Albert: crescimento da rede com o tempo e ligações preferenciais, permitem a construção de uma rede que tenha a distribuição dos graus de seus vértices seguindo uma lei de potência. A construção da rede, neste modelo, pode ser dada pelo seguinte esquema sequencial:

- i. Começando com um pequeno número de vértices  $m_0$ , a cada unidade de tempo, é adicionado um novo vértice com  $m \leq m_0$  arestas que conectam o novo vértice a  $m$  vértices diferentes já presentes na rede;
- ii. No momento da escolha dos vértices aos quais os vértices adicionados são conectados, assume-se certa probabilidade  $\Pi$  do novo vértice estar conectado ao vértice  $i$  que depende do grau  $k_i$  do vértice  $i$ . Deste modo a probabilidade pode ser escrita como

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (2.22)$$



Após um tempo  $t$ , este procedimento resultará uma rede de  $N = t + m_0$  vértices e  $A = mt$  arestas. Simulações numéricas indicam que a rede evolui em uma rede de estado invariante acerca da escala, de modo que as probabilidades de um vértice ter grau  $k$  é descrita pela distribuição de graus e segue uma lei de potência com  $\lambda > 1$  [12, 19]. O expoente de escala é independente de  $m$ , o único parâmetro do modelo.

### 2.2.8 Caracterização de redes complexas

Em qualquer estudo de uma rede complexa, que teve sua construção baseado em parâmetros encontrados na natureza ou na sociedade humana, algumas técnicas e quantificadores são necessários. Dados os modelos descritos anteriormente, é possível identificar em qual deles uma rede em estudo mais se adéqua. Para tanto, listam-se todas estas propriedades estruturais, que em conjunto, definem e caracterizam uma rede complexa:

1. Número total de vértices  $N$ : este é um parâmetro fundamental, pois além de ser um quantificador direto entre redes com mesma metodologia de construção, informa qual a confiabilidade de sua estrutura subjacente;
2. Número de arestas total da rede  $A$ : o número de arestas, assim como o número de vértices nos dá uma noção da proporção de conexões presentes na rede e é um parâmetro de comparação direto. Altos valores de  $A$ , para um mesmo  $N$ , indicam alta conectividade na rede e baixa tendência de organização; o inverso também é verdadeiro;
3. O grau médio  $\langle k \rangle$ : que nada mais é do que a média dos graus dos vértices que compõem a rede. Esta informação está associada, também, com a conectividade da rede;
4. A distribuição de graus  $p(k)$ : é a probabilidade de encontrarmos um vértice de grau  $k$  na rede complexa. Esse dado é muito importante, pois por meio dele é possível testar se a rede se adéqua ao modelo de rede livre de escala de Barabási-Albert dado pela equação 2.21 [20];
5. A distância de caminho médio  $\ell$ : seu cálculo é dado pela equação 2.19, e é uma das medidas utilizadas na identificação do processo de “small-world” na rede;
6. O diâmetro da rede, ou a maior geodésica  $\ell_{max}$ : esta é uma das características que também definem classes de grafos. Nos estudos de redes sociais, Milgram

[41] verificou que todas as pessoas estão conectadas por apenas outras 6 pessoas, os conhecidos de conhecidos; esta conclusão nos diz que o diâmetro da rede tem valor 6;

7. Coeficiente de aglomeração  $C_i$ : esse quantificador indica o quanto um vértice qualquer  $i$  está organizado em comunidades. Este cálculo está explicitado na equação 3.20 [18]. Como caracterizador da aglomeração geral na rede, basta tomarmos a média do coeficiente de aglomeração de todos os vértices  $\langle C_i \rangle$ ;
8. Identificação da estrutura hierárquica na rede: redes complexas podem apresentar comportamento modular, isto é, há grupos de nós que se organizam de modo hierárquico [42]. Para verificar essa propriedade em redes complexas, basta tomarmos o coeficiente de aglomeração médio por grau, ou seja:

$$\langle C_i(k) \rangle = \frac{\sum_{i=1}^n C_i}{n}. \quad (2.23)$$

onde  $C_i(k)$  é o vértice  $i$  de grau  $k$  e  $n$  o número de vértices com grau  $k$ . Então verificamos a dependência do tipo lei de potência de cada um destes coeficientes, de modo que

$$\langle C_i(k) \rangle \sim \frac{1}{k}. \quad (2.24)$$

9. Identificação da estrutura “small-world”: Para uma rede inicial, é possível calcular seu coeficiente de aglomeração da rede  $\langle C_i \rangle$  e a distância de caminho médio  $\ell$ . Tomando a distribuição de graus desta mesma rede, mas reconstruindo-a por meio da conectando seus vértices aleatoriamente, será obtido um grafo  $G$  aleatório, onde um novo coeficiente de aglomeração  $\langle C_i \rangle_{rand}$  e uma nova distância de caminho médio  $\ell_{rand}$  poderam ser calculados. Se compararmos estas quatro medidas, deve haver a seguinte relação

$$\ell_{rand} \approx \ell \text{ e } \langle C_i \rangle \gg \langle C_i \rangle_{rand}. \quad (2.25)$$

Se estas condições forem satisfeitas, a rede será do tipo “small-world”;

10. Assortatividade do grau: essa é propriedade da rede que reflete a correlação entre todos os pares de vértices por meio de seus graus. Este coeficiente de assortatividade de grau  $r$  varia de  $-1$  a  $1$ , de modo que valores positivos refletem a tendência de vértices de alto grau estar conectados a vértices de alto grau, e valores negativos refletem a tendência de vértices com alto grau conectados a vértices de baixo grau;
11. Medida de centralidade de intermediação (“betweenness”): essa medida indica a importância de um vértice na rede. Ela depende tanto do grau do vértice quanto do número de geodésicas que passam pelo vértice. Seja  $\sigma(i, j)$  o número de geodésicas entre os vértices  $i$  e  $j$ , e  $\sigma_l(i, j)$ , o número de geodésicas que saem de  $i$  e chegam a  $j$  passando necessariamente por  $l$ , então a centralidade de intermediação no vértice  $l$  é dada por

$$g_l = \frac{2}{(N-1)(N-2)} \sum_{i \neq j} \frac{\sigma_l(i, j)}{\sigma(i, j)}, \quad (2.26)$$

onde  $g_l$  é valor de centralidade [43];

12. Tamanho da componente gigante: o tamanho da componente gigante é o conjunto de vértices que possui uma alta densidade de arestas entre eles mesmos. O tamanho da componente gigante pode ser entendido como a máxima percolação na rede ou como a porção da rede que é conectada: isso significa que existe um caminho entre dois vértices quaisquer contidos na componente gigante [12]. O cálculo da componente gigante foi realizado por meio da função cluster da Linguagem Interpretativa R [24];
13. Robustez ou vulnerabilidade a ataques: removendo conjuntos de vértices com maior centralidade de intermediação, calculamos o tamanho da componente gigante. Se o decaimento for muito superior à escolha aleatória dos vértices, diz-se que a rede é vulnerável a ataques direcionados, se não parece diminuir em muito o tamanho da componente gigante, diz-se que é robusta a ataques direcionados. Para os ataques aleatórios, seguem-se as mesmas conclusões para cada caso;
14. Identificação de estruturas em comunidades: utilizando o “*Walktrap algorithm*” [45], também implementado em R, é possível identificar as

comunidades nas redes. Este algoritmo usa a caminhada aleatória na rede, e determina as partes mais densas na rede como a porção de vértices onde há menor variação no número de caminhantes por intervalo de tempo [25].

Ainda, todas as 11 primeiras propriedades, podem ser testadas independentemente dentro de cada comunidade encontrada pelo método descrito no item 13. Todas as ferramentas aqui descritas são uma revisão das principais formas de se caracterizar uma rede complexa em que a fenomenologia que modela esteja refletida em sua configuração. Nos capítulos subsequentes, estes procedimentos serão essenciais na caracterização de redes complexas.

## **2.3 Considerações finais do capítulo**

Neste capítulo foram revisados os elementos principais da teoria de grafo de forma a dar subsídio a teoria de redes complexas. Também foram introduzidos os conceitos fundamentais de redes complexas. Adicionalmente, foram tratados tanto aspectos históricos quanto o desenvolvimento da teoria. Analogamente, foram apresentados os principais modelos estudados atualmente: os modelos de redes randômico de Erdos e Renyi, de redes “small-world” de Watts e Strogatz e de redes livre de escala de Barabási e Albert. Finalmente foi possível listar as principais ferramentas utilizadas na caracterização de redes complexas que darão subsídio nos capítulos subsequentes.

Todos os estudos realizados neste capítulo tratam, em seu âmago, de sistemas em que a emergência tem um papel determinante. Isto no sentido de que a contemplação ideal do sistema seja dada pela modelagem em grafo, ou redes, do sistema real em estudo. Como no problema das Pontes de Königsberg, a solução veio apenas com a introdução do conceito de grafo que possibilitou a prova da ausência de um único caminho euleriano.

## CAPÍTULO 3

### Caminhada aleatória em redes complexas

A caminhada aleatória, por vezes denotada “o caminhar do bêbado”, é um modelo matemático formalizado sob a ideia intuitiva de que o estado de um sistema dependa de passos aleatórios em todas as direções. O exemplo mais comum é uma partícula que se difunde num meio homogêneo admitindo mudança de direção e posição, ambos aleatórios. Com o tempo, verificou-se que para essa fenomenologia funcionar, algumas condições devem ser satisfeitas, como: apenas um ponto de partida do movimento, a distância entre passos deve ser constante e a direção de um estado ao outro deve ser escolhida aleatoriamente.

Neste capítulo pretende-se abordar os conceitos mais usuais e básicos da teoria da caminhada aleatória nos casos em uma dimensão e em duas dimensões. Com isso, busca-se introduzir o formalismo da caminhada em redes complexas, que é o foco deste capítulo. Seja a finalidade gerar um quantificador algorítmico de transmissão de informação em redes complexas, no desfecho do capítulo será proposta uma nova forma de descrever estados gerais de sistemas aos quais possam ser modelados pela caminhada aleatória.

#### 3.1 Caminhada aleatória

##### 3.1.1 Caminhada aleatória: histórico e desenvolvimento

O problema da difusão num meio homogêneo foi abordado desde a época clássica, até os dias de hoje. A difusão aleatória congrega um conjunto de fenômenos observáveis, como a dispersão de fumaça numa sala isolada, uma substância dissolvida num líquido ou mesmo o movimento de enxames de insetos. E a esta noção intuitiva de difusão, deu-se o nome de caminhada aleatória, termo originalmente proposto por Karl Pearson em 1905. No entanto, a primeira tentativa de descrição do fenômeno ocorre em 1827 com Robert Brown, botânico e inglês, que publicou suas observações do comportamento de grãos de pólen em meio aquoso. Em seu trabalho, Brown verificava que quando em solução aquosa, os grãos de pólen tinham

movimentos irregulares, tornando impossível prever a mudança de posição e direção no tempo, descrevendo dependência aleatória na movimentação.

Inspirado pelos trabalhos de Brown, Pascal, Fermat e Bernoulli, Pearson concretizou a teoria de caminhada aleatória que viria a ser interpretada como um processo estocástico. Em uma Carta à *Nature*, Pearson apresentou um modelo simples para descrever o comportamento de enxames de mosquitos numa floresta [26], de modo que a cada incremento de tempo, um único mosquito se move a uma distância fixa em uma direção aleatória. Após um tempo, e considerando um grande número de mosquitos, o enxame não se moveria além de certo limite.

Simultaneamente à Pearson, Louis Bachelier também propôs uma teoria de caminhada aleatória publicada em 1900 sob o título “La Théorie de La Speculation”, onde aplicou a caminhada aleatória como modelos para séries temporais em economia, permanecendo até hoje como uma das bases da teoria moderna de finanças. Foi também com Bachelier que foi realizada a primeira relação entre caminhadas aleatórias discretas e equações de difusão no meio contínuo [27].

Em um trabalho muito análogo ao de Pearson, Albert Einstein publica seu trabalho sobre o movimento browniano em 1905, ao qual estudou a difusão de partículas cujas trajetórias são governadas por colisões aleatórias [28]. Seu propósito era obter uma expressão para calcular o número de Avogadro por meio de quantificadores do processo de difusão, por exemplo, em moléculas na atmosfera.

É notável a generalidade de aplicação da caminhada aleatória como modelo de descrição de sistemas reais. Há exemplos da biologia à física, da matemática pura à economia. Porém, nesta seção, pretende-se explorar algumas simples formalidades desta teoria.

### 3.1.2 Caminhada aleatória em uma dimensão

Consideremos uma reta onde se encontram inscritas posições equidistantes de tamanho  $l$ . E um caminhante que caminhe  $N$  passos de tamanho  $l$ , de modo que em cada intervalo de tempo discreto sua posição coincida com as posições equidistantes. Admitindo-se, também, que  $p$  é a probabilidade de dar um passo à direita e  $q$  à esquerda, temos que

$$p + q = 1. \tag{3.1}$$

de modo que a probabilidade seja conservada [29]. O número de passos  $N$  pode ser decomposto em passo dados à direita  $n_1$  e à esquerda  $n_2$  de modo que

$$n_1 + n_2 = N. \quad (3.2)$$

Partindo do ponto de vista de dar passos à direita, haverá uma combinação de  $\binom{N}{n_1} = \binom{n_1+n_2}{n_1}$  maneiras de tomar  $n_1$  passos à direita e  $n_2$  à esquerda. Onde uma sequência particular de passos, constituídas pelas probabilidades  $p^{n_1}q^{n_2}$ , caracteriza uma distribuição binomial do tipo:

$$P(n_1) = \frac{(n_1 + n_2)!}{n_1! n_2!} p^{n_1} q^{n_2} \therefore$$

$$P(n_1) = \frac{N!}{n_1! (N - n_1)!} p^{n_1} q^{n_2}. \quad (3.3)$$

Seja o número esperado de passos  $n_1$  para a direita dado por

$$\langle n_1 \rangle = pN. \quad (3.4)$$

então

$$\langle n_2 \rangle = qN,$$

$$\langle n_2 \rangle = N - \langle n_1 \rangle,$$

$$\langle n_2 \rangle = N(1 - p). \quad (3.5)$$

Analogamente, o valor esperado do segundo momento ou a variância é dado por

$$\sigma_{n_1}^2 = \langle n_1^2 \rangle - \langle n_1 \rangle^2 \equiv Npq. \quad (3.6)$$

E o desvio padrão:

$$\sigma_{n_1} = \sqrt{Npq} \quad (3.7)$$

Também é possível descrever a distância percorrida pelo caminhante na reta após certo número de passos:

$$d_N = (n_1 - n_2)\ell \equiv (2n_1 - N)\ell. \quad (3.8)$$

onde  $\ell$  é o tamanho de cada passo. Em um caso especial, onde  $p = 0,5$ , a probabilidade  $P_N(d)$  de um caminhante percorrer uma distância  $d$  após  $N$  passos pode ser escrita como

$$P_N(d) = \frac{1}{2^N} \binom{N}{\frac{d+N}{2}}, \quad (3.9)$$

em que os momentos de valores esperados podem ser dados por uma única expressão genérica

$$\mu_p = \sum_{d=-N, -(N-2), \dots, N} d^p P_N. \quad (3.10)$$

E o valor esperado da distância total percorrida após  $N$  passos pode ser escrita como

$$\langle d_N \rangle = \sum_{d=-N, -(N-2), \dots, N} |d| P_N(d). \quad (3.11)$$

Tornando claras algumas formalidades do modelo da caminhada aleatória em uma dimensão, passe-se a algumas considerações da caminhada para duas dimensões.

### 3.1.3 Caminhada aleatória em duas dimensões

Citando novamente o trabalho de Pearson e seu modelo que descreve o movimento de um enxame de mosquitos, Lord Rayleigh responde Pearson, defendendo que este problema de difusão de partículas aleatórias, em mais de uma dimensão, já havia sido resolvido em 1880.



Em um modelo de propagação de ondas sonoras em um material, de forma análoga ao movimento de mosquitos, pode ser idealizado como a soma dos vetores de onda com  $k$  constantes e fases randômicas. E com o objetivo de encontrar a função densidade de probabilidade de uma onda sonora após muitas interações com um material (espalhadores), Lord Rayleigh encontrou a dependência da probabilidade  $P_N(R)dR$  de atravessar a distância entre  $R$  e  $R + dR$  em  $N$  iterações ou passos. Para um caso discreto e  $N \rightarrow \infty$ , tem-se

$$P_N(R) \sim \frac{2R}{N} e^{R^2/N}. \quad (3.12)$$

De modo que o valor esperado da distância percorrida é proporcional à  $\langle R^2 \rangle \propto N$ . Para o caso do estudo de mosquitos de Pearson em duas dimensões, tem-se que  $d = 2$ ,  $p(r) = \delta(r - a)$  e  $\langle r^2 \rangle = a^2$ , logo

$$P_N(R) = \frac{2R}{a^2 N} e^{R^2/a^2 N}, \quad (3.13)$$

em que  $a$  é a distância percorrida pelo mosquito a cada incremento de tempo. Interessantemente, as relações 3.12 e 3.13 são concordantes, pois no modelo de Rayleigh foi considerado  $a = 1$ .

### 3.1.4 Caminhada aleatória em redes complexas

Seja o interesse geral deste manuscrito estudar a caminhada aleatória especialmente em redes representadas por grafos direcionados. Consideremos um dígrafo  $G = (V, A)$ , de modo que para cada aresta direcionada (*i.e.*, arco)  $r = \langle \overline{u, v} \rangle$ , existe um peso associado  $0 < m_{\overline{uv}} \leq 1$ . O valor  $m_{\overline{uv}}$  denota o peso da aresta de  $u$  alcançar  $v$  em um passo. Com isso, fica restringido que a soma das probabilidades de peso deve ser 1. A probabilidade de peso é uma grandeza arbitrária, mas pode estar associada à natureza da interação entre os vértices. Por exemplo, a força de ligações iônica ou covalente entre elementos numa rede cristalina, o aproveitamento trófico de uma espécie ao se alimentar de outra, a frequência com que duas pessoas interagem, etc.

Se denotarmos  $M_G$  a matriz de probabilidade de pesos com entradas  $m_{\overline{uv}}$ , então  $M_G$  será uma cadeia de Markov sobre o grafo  $G$ . Isso implica que o conhecimento dos estados

anteriores é irrelevante na determinação da probabilidade de estados futuros, ou seja, o próximo estado do sistema depende totalmente do estado atual. É importante salientar que as cadeias de Markov tratadas neste manuscrito são de ordem- $m$ , para  $m = 1$ ; isso implica que o próximo estado do sistema depende apenas nos  $m$  estados anteriores [31]. Formalmente:

$$Pr[X_{t+1} = v | X_t = u, X_{t-1} = u_1, \dots, X_0 = u_t] = Pr[X_{t+1} = v | X_t = u], \quad (3.14)$$

em que  $Pr[X, Y]$  representa a probabilidade de transitar de um estado  $X$  ao estado  $Y$ . A distribuição de probabilidades sobre um grafo  $G$  em qualquer tempo, dá a probabilidade de cada vértice atravessar para aquele vértice. Também, a distribuição de probabilidade é um vetor unidimensional de tamanho  $n$ . Dado que  $P^t$  denota a distribuição de probabilidade após  $t$  tempos, tem-se

$$P^t = (Pr_i, Pr_j, \dots, Pr_n), \quad (3.15)$$

onde  $P_i$  denota a probabilidade de transição no vértice  $i$ . Logo:

$$\begin{aligned} P^{t+1} &= P^t M_G, \\ P^t &= P^0 M_G^t, \end{aligned} \quad (3.16)$$

onde  $P^0$  é uma distribuição inicial. Com isso é possível escrever formalmente:

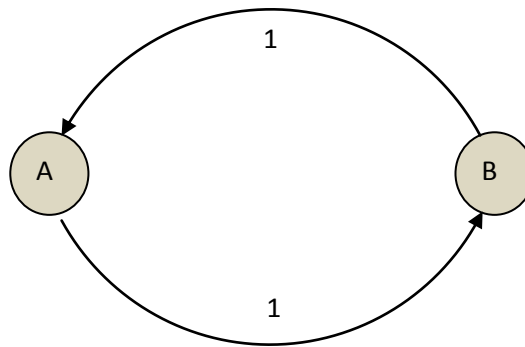
**Definição 3.1.** *Uma distribuição  $\Pi$  será estacionária se  $\Pi M_G = \Pi$ . Isto simplesmente significa que uma vez alcançado  $\Pi$ , a distribuição não mais mudará.*

Alguns pontos importantes derivam desta definição:

- I. A cadeia de Markov admite alguma distribuição estacionária?
- II. Se admitir, quantos passos ou unidades de tempo são necessários para alcançar essa distribuição?
- III. Há apenas uma distribuição estacionária?

No geral, pode não ser possível alcançar uma distribuição estacionária mesmo que ela exista dada uma cadeia de Markov conhecida, pois a alcançabilidade de uma distribuição estacionária depende da distribuição inicial [31]. Por exemplo, se considerarmos uma cadeia de Markov onde há dois estados, como ilustrado na figura 4.1; há uma distribuição estacionária  $\left(\frac{1}{2}, \frac{1}{2}\right)$  a qual não é obtida se iniciarmos as com distribuições iniciais  $(1,0)$  e  $(0,1)$ .

Figura 3.1 Exemplo de uma cadeia de Markov.



Fonte: autor

A única forma de se obter a distribuição estacionária  $\left(\frac{1}{2}, \frac{1}{2}\right)$  é quando o estado inicial coincide com ela mesma [31]. Dada essa dependência entre a distribuição estacionária e a distribuição inicial, é necessário estudar as condições que implicam nesta dependência. Para tanto, tem-se:

**Definição 3.2.** *Seja  $u$  e  $v$  dois vértices. É enunciado que  $u$  e  $v$  se comunicam se  $\exists n > 0, m > 0$  tal que*

$$M_{uv}^n > 0, M_{vu}^m > 0,$$

onde  $M_{uv}^n$  denota a probabilidade de uma cadeia de passos começar em  $u$  e estar em  $v$  após  $n$  passos.

Ainda.

**Definição 3.3.** *Uma cadeia de Markov sobre um grafo  $G$  é irredutível se cada estado estiver comunicado a todos os outros estados.*

É importante notar que uma cadeia de Markov em  $G$  será irredutível se  $G$  for um grafo conectado, ou seja, se for possível transitar de qualquer vértice ao outro em um tempo finito (ver Capítulo 2). Com isso fica fácil ver que a irredutibilidade é a condição mínima para alcançar uma distribuição estacionária única. De modo que se essa propriedade não for satisfeita, não será possível alcançar alguns dos vértices do grafo dada uma distribuição inicial. Com isso, é necessário determinar a condição que elimine a impossibilidade de alcançar uma distribuição estacionária, como ilustrado na Figura 3.1. Para tanto:

**Definição 3.4.** *Um vértice  $v$  será periódico se houver um  $\Delta > 1$  tal que  $Pr[X_{t+s} = v | X_t = v] = 0$  a não ser que  $s$  seja divisível por  $\Delta$ .*

Isso significa que iniciando uma caminhada em  $v$ , e após  $\Delta$  passos, será ou não possível retornar a  $v$ ; porquanto será impossível voltar a  $v$  se  $s$  não for um múltiplo de  $\Delta$ . Analogamente, uma cadeia de Markov será aperiódica se todas suas distribuições forem, também, aperiódicas.

### 3.1.5 Teorema fundamental da cadeia de Markov

Para toda cadeia de Markov finita, irredutível e aperiódica, haverá uma distribuição estacionária  $\Pi$  que pode ser obtida de qualquer distribuição inicial. Considerando que a distribuição estacionária seja denotada por

$$\Pi = (\Pi_1, \Pi_2, \dots, \Pi_n). \quad (3.17)$$

De modo que

$$\Pi_i = \frac{1}{h_{ii}}. \quad (3.18)$$

O termo  $h_{ii}$  é o tempo esperado para que uma caminhante saia do vértice  $i$  e retorne a ele. É importante notar que se uma distribuição estacionária existe para um dado grafo, então deve ser possível a obter pela distribuição inicial.

No geral, considerando novamente  $u$  e  $v$  dois vértices de  $G$ . Seja  $M_{uv}^t$  a probabilidade de alcançar  $v$  de  $u$  em exatamente  $t$  passos. Então o tempo esperado de ir de  $u$  a  $v$  será

$$h_{uv} = \sum_{t \geq 1} t \cdot M_{uv}^t . \quad (3.19)$$

Com isso,  $h_{uu}$  será o tempo de retorno a  $u$ . Estes são os aspectos mais básicos sobre processos Markovianos que simplificam a interpretação de caminhada aleatória em rede complexa.

## 3.2 Considerações finais do capítulo

Neste capítulo foram definidas as formas mais usuais e simplificadas na caminhada aleatória. Foram dados os conceitos que fundamentam essa fenomenologia e como se tornou um modelo para descrever processos naturais em diversos campos de pesquisa. Em uma e duas dimensões, foram mostrados os quantificadores principais acerca de cada um. Então foi estudada a caminhada em redes complexas como uma cadeia de Markov, podendo ou não assumir uma distribuição estacionária. As noções desenvolvidas nesse capítulo irão servir de base para a construção de um modelo mais específico o aprofundado no capítulo 5.

## **CAPÍTULO 4**

### **A rede mitológica da Odisseia de Homero**

A natureza intrigante das narrativas clássicas de Homero tem fascinado a cultura ocidental desde sempre, contribuindo na filosofia, história, mitologia e obviamente na literatura. Porém, o que torna tais narrativas tão intrigantes? Em um primeiro momento, é evidente o apelo literal e a estética textual apresentada em cada página dos cânticos da Odisseia e das rapsódias da Ilíada. Em segundo lugar, é possível captar subjetivamente um aspecto enviesado nos conteúdos das histórias narradas por Homero, parecendo variar do histórico-real ao mitológico-ficcional. Ainda, em 2004, Korfman relata alguns achados arqueológicos que deram suporte a alguns dos eventos descritos na Ilíada, como uma grande guerra entre os povos que viviam onde hoje é a atual Turquia e a Grécia clássica. Considerando todas estas nuances acerca da historicidade dos conteúdos presentes nas obras de Homero, neste capítulo iremos propor uma análise da rede social contida na Odisseia, de modo a dar algumas pistas no intento de conhecer sua historicidade.

#### **4.1 Mitologia comparativa e redes mitológicas**

##### **4.1.1 Contextualização do uso redes complexas em mitologia comparativa**

A mudança de paradigma, do reducionismo ao holismo, é frequente nos dias atuais de modo a aumentar o interesse interdisciplinar dentre cientistas [33]. Este processo pode ser realizado à medida que os conceitos fundamentais da teoria de redes complexas são aplicados a problemas que surgem em diversas áreas do conhecimento, como: redes sociais, comunicação, economia, bolsa de valores, ciências computacionais, Internet, WWW, sistemas de transporte, distribuição da rede elétrica, biologia molecular, ecologia de comunidades, neurociência, linguística, meteorologia e assim por diante. No entanto, à medida que os estudos são realizados utilizando redes complexas para descrever fenômenos, padrões estruturais emergem, tornando possível identificar classes de redes [13, 21, 25]. Considerando

esta possibilidade, podemos construir certa taxionomia de redes pela simples comparação de sua composição estrutural, ou “topologia” [35]. Em conformidade com esse pensamento, a noção de universalidade também é um conceito-chave na unificação de vários tipos de redes em grupos característicos [12].

Baseados neste conceito de universalidade, Carron e Kenna [36] propõem um método de análise para discriminar uma narrativa qualquer entre real ou ficcional; faz-se isso utilizando a rede social ao qual a narrativa se refere. Especificamente, estes autores analisaram três narrativas clássicas de historicidade incerta: Beowulf, Iliada e Táin Bó Cuailnge. A partir destas três narrativas, foram construídas redes sociais onde os vértices representam as personagens e as arestas sua interação social na estória. Como resultado, foi gerado um tipo de rede social denominada pelos autores de redes mitológicas [36]. Em contrapartida, as redes sociais reais, em conjunto, têm algumas propriedades bem conhecidas, como: são frequentemente “small-world” [18, 34], têm organização hierárquica [22, 39, 40], são altamente aglomeradas, tem assortatividade de grau positivas [37, 38] e são redes livre de escala [20, 34]. Além destas características, espera-se que redes sociais reais tenham distribuição de grau to tipo lei de potência, apresentem componente gigante maior que 90%, sejam vulneráveis a ataques direcionados e robustos sob ataques aleatórios [41].

Em outro extremo, outra configuração é possível à redes sociais colhidas de narrativas: as redes sociais fictícias. De forma análoga às redes sociais reais, estas redes também podem ser caracterizadas: são “small-world”, possuem estruturas hierarquizadas, têm distribuição de grau do tipo exponencial, não são redes livre de escala, suas componente gigante são menores que 90%, não apresentam assortatividade no grau e são robustas tanto para ataques direcionados quanto ataques aleatórios. Apesar de algumas propriedades serem semelhantes às redes sociais reais, as fictícias quase sempre apresentam seu caráter distinto com uma análise profunda, principalmente acerca dos ataques aos vértices mais importantes [36, 40, 42]. Também é notório que a distribuição de grau e um fator diferenciador entre as redes sociais reais e fictícias; uma sendo uma lei de potência e outra sendo uma lei exponencial, respectivamente.

Com estas propriedades específicas, Carron e Kenna verificaram que a *Ilíada* de Homero, das obras estudadas, foi aquela que mais se aproximou de uma rede social real, enquanto que para Beowulf e Táin, fora preciso fazer algumas modificações para mostrar o caráter real de suas redes sociais. Embora realizadas manipulações, estes autores sintetizaram um meio para analisar contos, mitos e narrativas clássicas. Tal síntese pode ser utilizada para identificar a estrutura social da época que a narrativa trata de modo a tornar uma poderosa

ferramenta à mitologia comparativa. Adicionalmente, é importante citar a obra de Campbell, “The Hero with a thousand faces”; onde descreve que apesar de os mitos estarem presentes em todas as civilizações e ter origens diversas, sua estrutura essencial é quase sempre a mesma. A esta coincidência, Campbell denomina o monomito [43]. Associando o conceito do monomito de Campbell e os resultados de Carron e Kenna, é possível especular que um conto clássico como a Ilíada, é uma mistura de conteúdo mitológico associado com alguma informação histórica dos povos da época. Isso pode ser realizado dessa forma para que a obra, como um todo, seja mais atraente literalmente, possibilitando maior penetração nas gerações subsequentes.

Inspirado por estas possibilidades especulativas, propomos a análise da rede social do épico homérico, A Odisseia [44]. Buscando algum significado em termos da constituição estrutural de sua rede social, pretende-se identificar se a rede da Odisseia se aproxima mais do padrão real ou do ficcional, considerando as implicações para a sua mitologia comparativa. Além de analisar uma obra nova, a Odisseia, propomos a análise de comunidades na rede utilizando um algoritmo de caminhada aleatória [45]. Ainda, com as comunidades, pretende-se realizar uma análise da constituição de suas personagens contextualizando as comunidades na estória.

A novidade e ênfase deste estudo em particular consistem na forma de interpretar os resultados acerca da composição estrutural em termos da composição das personagens nas comunidades. Espera-se que disso resulte uma explicação plausível da estrutura como um todo em termos das relações especiais entre as personagens e seu contexto na estória. Além da descrição estatística da rede como um todo, é fundamental entender a forma com que a rede possibilite encontrar fenômenos reais, como agrupamentos de personagens em torno de eventos, famílias, nações e grupos.

#### **4.1.2 Descrição literária do poema da Odisseia**

Juntamente com a Ilíada, a Odisseia de Homero exprime com força e beleza as maravilhas da remota civilização grega. Estes clássicos datam por volta do século VIII A.C., após o desenvolvimento do sistema de escrita herdada do alfabeto fenício [47, 48]. É sabido que a Odisseia carrega os ecos da Guerra de Tróia narrada principalmente na Ilíada. Lembrando-se do trabalho de Carron e Kenna, e suas análises de estórias mitológicas, a rede mitológica das personagens da Ilíada foi a mais similar às redes sociais reais encontradas atualmente. Adicionalmente, algumas evidências arqueológicas encontradas na região da



Anatólia dá suporte à historicidade de alguns conflitos que ocorreram por volta do século XII A.C., associados diretamente ao conflito entre gregos e troianos daquele tempo [49, 50].

Como um *continuum* da *Ilíada*, a *Odisseia* narra as desventuras do herói Odisseu em sua luta de volta para seu lar. Essa saga toma 10 anos para que Odisseu finalmente alcance Ithaca, sua terra natal, após 10 anos de guerra em Tróia, somando 20 anos de tormenta. O épico é um poema composto por 24 cânticos de versos hexâmeros. É importante salientar que a obra tem ordem temporal inversa no curso dos eventos: começando com o desfecho, ou seja, a assembleia dos deuses quando Zeus decide que Odisseu deve finalmente retornar ao lar. O texto é estruturado em quatro partes principais: o primeiro (cânticos I a IV), intitulado “A assembleia dos deuses”; o segundo (cânticos V a VIII), “A nova assembleia dos deuses”; o terceiro (cânticos IX a XII), “O colóquio de Odisseu”; e o quarto (cânticos XIII a XXIV), “De volta para casa”.

A obra-prima que constitui a *Odisseia* é tomada como um conjunto de aventuras frequentemente consideradas mais profundas e complexas que *Ilíada*; pois há muitos aspectos epopeicos que aproximam o leitor à verdadeira natureza humana, enquanto que na *Ilíada* predomina os aspectos heroicos, legendários e divinos em suas personagens. Apesar desta disparidade, é consenso que a *Odisseia* completa o contexto iniciado na *Ilíada*, contemplando grande parte da bagagem cultural da Grécia Antiga. É óbvia, também, a genialidade de Homero, dada a influência destas duas obras para a cultura ocidental, tornando o imaginário coletivo muito mais rico e propenso a identidade cultural do homem atual [44].

## 4.2 Análise da rede mitológica da *Odisseia*

### 4.2.1 Coleta de dados textuais

O poema da *Odisseia* é composto por um vocabulário pouco usual tornando a leitura e interpretação do conteúdo um empreendimento homérico. Muitas vezes é complicado identificar a ação das personagens apenas por seus nomes, pois muitas vezes estão associadas a títulos alegóricos ou a adjetivos de seus feitos heroicos. Por exemplo: “Canta, Ò Musa, o varão que astucioso...”, o varão astucioso é uma alusão ao personagem Odisseu; “Então disse Atena: Pai Filho de Chrono, Reis dos reis...”, a deusa Atena está se referindo a seu pai, Zeus. Dada essas dificuldades, é necessário adotar uma forma de analisar o conteúdo textual profundamente, de modo a identificar a quem o interlocutor se refere. Este viés gera uma

impossibilidade da extração direta do texto, como, por exemplo, na mineração textual [52]. Com isso, propõe-se um método de construção da rede baseada na interpretação textual do poema.

Como interpretação pode ter um viés muito subjetivo, e é primado por um processo menos arbitrário possível, para tanto se define alguns critérios normativos na extração textual. Como em redes sociais derivadas de narrativas, os vértices representam personagens (convencionado como entidades nomeadas) e arestas representam pelo menos uma forma de interação entre personagens. Daqui, faz-se necessário definir tais interações por meio dos seguintes itens:

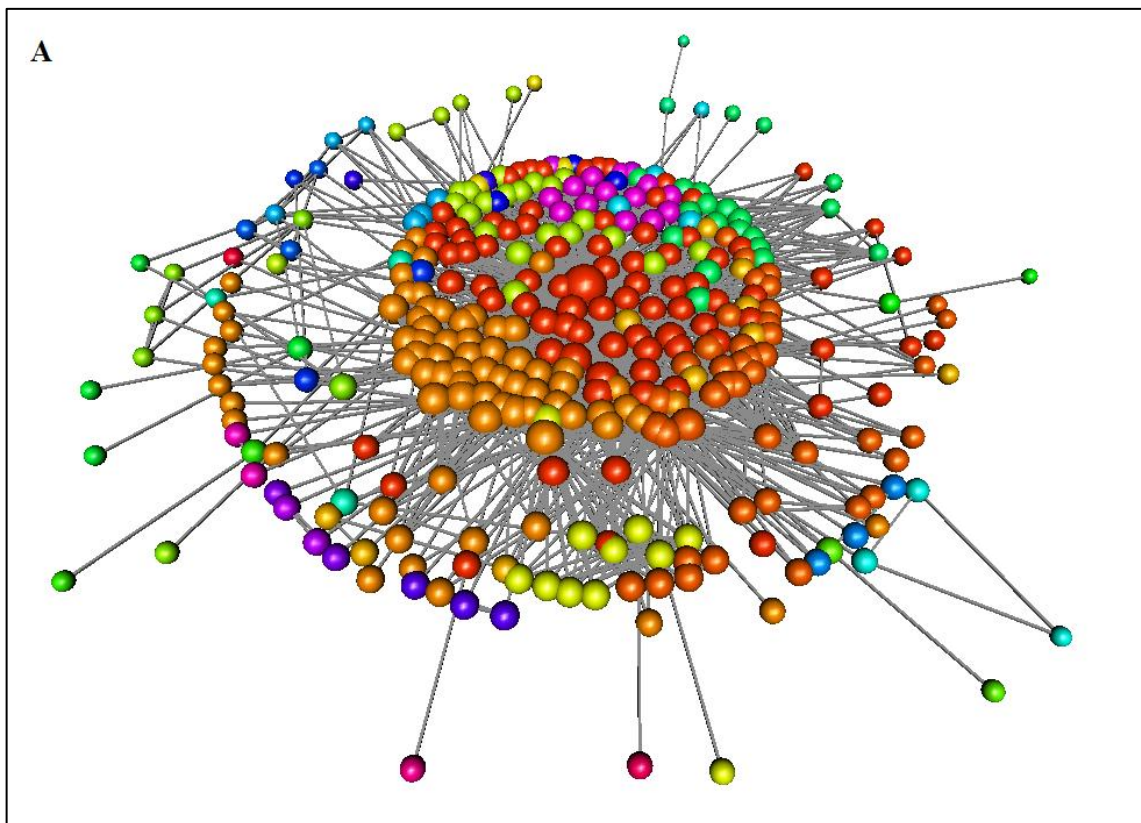
1. Personagens que estiverem no mesmo local no mesmo instante, demonstrando participar da ação: critério de compartilhamento de evento;
2. Personagens que interajam diretamente: critério de interação direta;
3. Personagens que citam outra à terceiros. Esse critério deve prevalecer se, no ato da citação, o interlocutor especificar um mínimo de conhecimento acerca de seu subjecto: critério de interação indireta.

Para evitar interpretações equívocas das relações sociais contidas no poema, foram utilizadas diferentes traduções e edições da Odisseia de Homero [44, 47, 48, 51]. Como se esperava, não há diferenças significativas entre os conteúdos das traduções e edições, de modo que não há problemas na criação da rede quanto a este aspecto. Contudo, na análise cuidadosa do texto, foi possível identificar 342 personagens ligados socialmente por 1747 interações (Figura 4.1). É importante enfatizar que a rede é limitada em termos da sociedade da época que se passou a estória, logo, ela é uma representação de parte da sociedade. A análise estrutural de redes está descrita na seção 2.2.8 do segundo capítulo deste manuscrito; esta análise irá se limitar a enquadrar a rede em real ou fictícia, ou seja, uma análise indutiva.

Adversamente, neste trabalho é proposta uma análise dedutiva baseada no conceito de comunidade e sua composição de personagens. O conceito de comunidade não é definido univocamente, pois há muitas variações no formalismo e nos formais e métodos para a definir [41]. No geral, ou as definições são muito restritivas ou o método é computacionalmente ineficiente. No entanto um consenso pode ser encontrado: se se considera uma partição  $P = \{C_1, \dots, C_k\}$  de vértices de um grafo  $G$  como o conjunto que constitui a comunidade, espera-se que a densidade de arestas internas seja muito superior as arestas entre outras

regiões da rede. Para processar essa identificação de comunidades por esse critério, introduz-se o algoritmo “walk trap”, que utiliza caminhadas aleatórias no grafo de modo a gerar quantificadores de tendência de os caminhantes se manterem “presos” a um conjunto definido de vértices por um intervalo de tempo maior. A este conjunto dá-se o nome de comunidade do grafo [45]. Com a identificação das comunidades, repete-se a análise de propriedades estruturais para as mais relevantes ( $V > 10$ ), buscando pelas estruturas significativas do ponto de vista da teoria de redes complexas.

Figura 4.1 Rede mitológica da Odisseia. A coloração dos vértices é a indicação da comunidade que as personagens pertencem. O tamanho do vértice é diretamente proporcional à centralidade de intermediação.



Fonte: autor.

#### 4.2.2 Análise indutiva

O sumário das propriedades estruturais encontradas na rede mitológica da Odisseia pode ser encontrado na tabela 4.1 juntamente com os dados compilados das obras analisadas por Carron e Kenna [36].

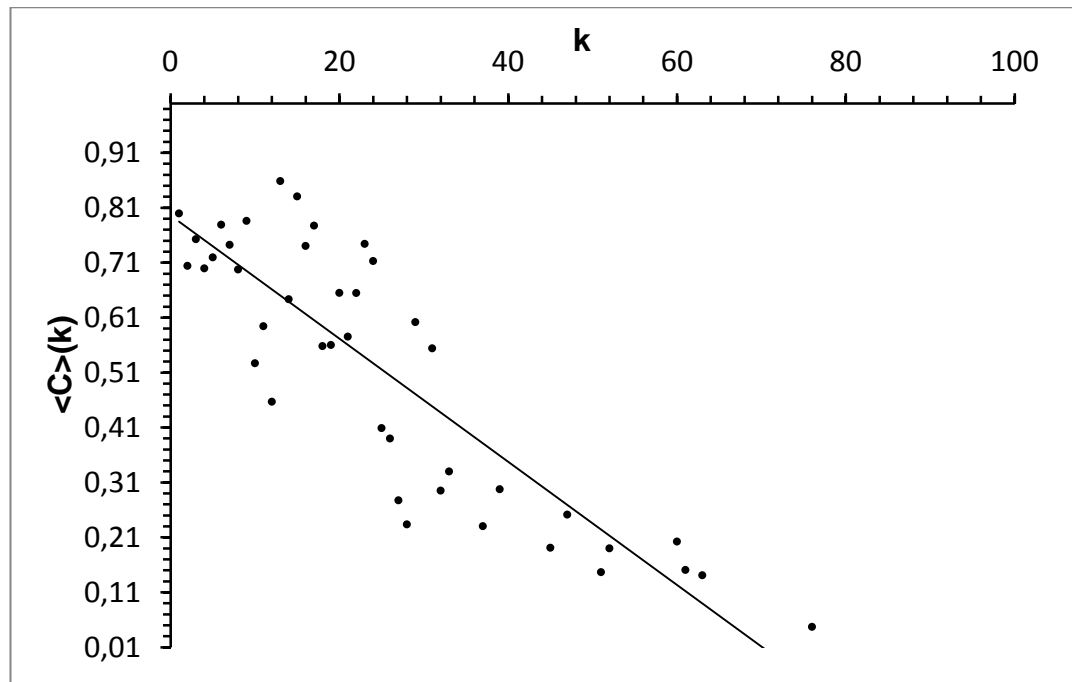
Tabela 4.1 Sumário das propriedades estruturais de diversas redes mitológicas: número de vértices ( $V$ ), número de arestas ( $A$ ), distância de caminho médio ( $\ell$ ), diâmetro ( $\ell_{max}$ ), coeficiente de aglomeração ( $C$ ), tamanho da componente gigante ( $G_c$ ) e assortatividade do grau ( $r$ ). Odisseia\*, Beowulf\* e Táin\* são as redes originais excetuadas as personagens mitológicas.

| Rede             | $V$ | $A$  | $\langle k \rangle$ | $\ell$ | $\ell_{rand}$ | $\ell_{max}$ | $C$  | $C_{rand}$ | $G_c$       | $r$   |
|------------------|-----|------|---------------------|--------|---------------|--------------|------|------------|-------------|-------|
| <i>Odisseia</i>  | 342 | 1747 | 10,22               | 2,58   | 2,47          | 6            | 0,27 | 0,11       | 342 (100%)  | -0,15 |
| <i>Odisseia*</i> | 318 | 1129 | 7,10                | 4,08   | 3,10          | 11           | 0,54 | 0,06       | 274 (86%)   | 0,09  |
| <i>Iliada</i>    | 716 | -    | 7,40                | 3,54   | 3,28          | 11           | 0,57 | 0,01       | 707 (98.7%) | -0,08 |
| <i>Beowulf</i>   | 74  | -    | 4,45                | 2,37   | 2,88          | 6            | 0,69 | 0,06       | 50 (67.5%)  | -0,10 |
| <i>Táin</i>      | 404 | -    | 2,76                | 2,76   | 3,32          | 7            | 0,82 | 0,02       | 398 (98.5%) | -0,33 |
| <i>Beowulf*</i>  | 67  | -    | 3,49                | 2,83   | 3,36          | 7            | 0,68 | 0,05       | 43 (64.2%)  | 0,01  |
| <i>Táin*</i>     | 324 | -    | 3,71                | 3,88   | 4,41          | 8            | 0,69 | 0,01       | 201 (62%)   | 0,04  |

Fonte: O autor (para Odisseia) e Carron & Kenna (2012) referência [36] (para Iliada, Beowulf e Táin).

Foi observado que a rede mitológica da Odisseia tem distância de caminho médio similar à distância de caminho médio de uma rede construída aleatoriamente ( $\ell \approx \ell_{rand}$ ). O coeficiente de aglomeração é maior quando comparado com o coeficiente de aglomeração de uma rede aleatória ( $C > C_{rand}$ ). Estas duas condições indicam que a rede tem algumas das características de “small world” como discutidas no capítulo 2. A organização hierárquica da rede está ilustrada na figura 4.2, onde o coeficiente de aglomeração médio por grau é graficado contra o grau,  $\bar{C}(k)$  vs.  $k$ . É possível verificar que os vértices de menor grau apresentam maior coeficiente de aglomeração em relação aos de maior grau, seguindo aproximadamente uma regra linear. Interpreta-se que vértices de alto grau integram pequenas comunidades gerando unificação da rede como um todo. Isso indica, que a rede não possui caráter hierarquizado, pois para tanto é preciso que a regra entre  $\bar{C}(k)$  e  $k$  fosse uma lei de potência.

Figura 4.2 Gráfico em escala log-log que demonstra a organização hierárquica da rede mitológica da Odisseia. São representados os coeficientes de aglomeração médios por grau contra os graus. A linha cinza pontilhada representa a lei de potência  $1/k$  e a linha preta é a regressão linear para a lei de potência dos dados coletados da rede (pontos pretos).



Fonte: autor.

O tamanho da componente gigante, observada para a rede mitológica da Odisseia, contém todos os seus vértices. Este resultado sugere duas possíveis interpretações: ou o fenômeno da componente gigante não ocorre nesta rede ou os dados acerca da rede não são suficientes de modo a apresentar esse fenômeno em sua totalidade. A estas alusões, é necessária uma análise mais aprofundada dada pelos ataques direcionado e aleatório como resposta no tamanho da componente gigante. Os dados de ataques estão organizados na tabela 4.2. As metodologias de ataques estão descritos na seção 2.2.8 do capítulo 2, item 13.

Tabela 4.2 Ataques direcionados e aleatórios como respostas na componente gigante  $G_c$ , em termos do valor absoluto e relativo.

| Ataque direcionado |             | Ataque aleatório |             |
|--------------------|-------------|------------------|-------------|
| $G_c$              |             | $G_c$            |             |
| Sem ataque         | 342 (100%)  | Sem ataque       | 342 (100%)  |
| 5%                 | 274 (79,6%) | 5%               | 332 (93,6%) |
| 10%                | 188 (54,6%) | 10%              | 309 (89,9%) |
| 15%                | 163 (47,3%) | 15%              | 282 (81,9%) |
| 20%                | 121 (35,1%) | 20%              | 273(79,3%)  |
| 25%                | 41 (11,9%)  | 25%              | 248 (72%)   |

Fonte: autor.

Os ataques direcionados demonstraram que a estrutura da rede, em termos de sua componente gigante, depende muito de suas personagens centrais, ou seja, com maior centralidade de grau (“betweenness”). No entanto, se são removidos vértices independentemente de sua centralidade, o tamanho da componente gigante não cai tão abruptamente. Logo, é possível dizer que a rede é vulnerável a ataque direcionado e robusto a ataque aleatório, característica fundamental em redes sociais reais [23, 37, 38]. É notável que esta propriedade é extensivamente explorada nos modelos de rede livre de escala, onde há uma tendência de haver poucos nós muito conectados e muitos nós pouco conectados [12, 20, 41]. Para testar o caráter livre de escala da rede, atenta-se para a distribuição de graus ilustrado na figura 4.3. Como parte da distribuição segue uma lei de potência para  $\gamma = 1,2 \pm 0,1$  (com  $\chi^2/gl = 0,06$ ), isso indica que a rede é livre de escala. Neste ponto, é importante fazer uma nota sobre o ajuste da distribuição de graus: a maior parte dos autores que tratam as distribuições de graus de redes sociais, fazem o ajuste da lei de potência apenas para a “cabeça” da distribuição, defendendo que esta é a porção mais relevante da distribuição [23, 36-38]. No entanto, propomos um outro tipo de ajuste, a Pareto Generalizada [106, 107]. Este ajuste é dado pela seguinte função de distribuição

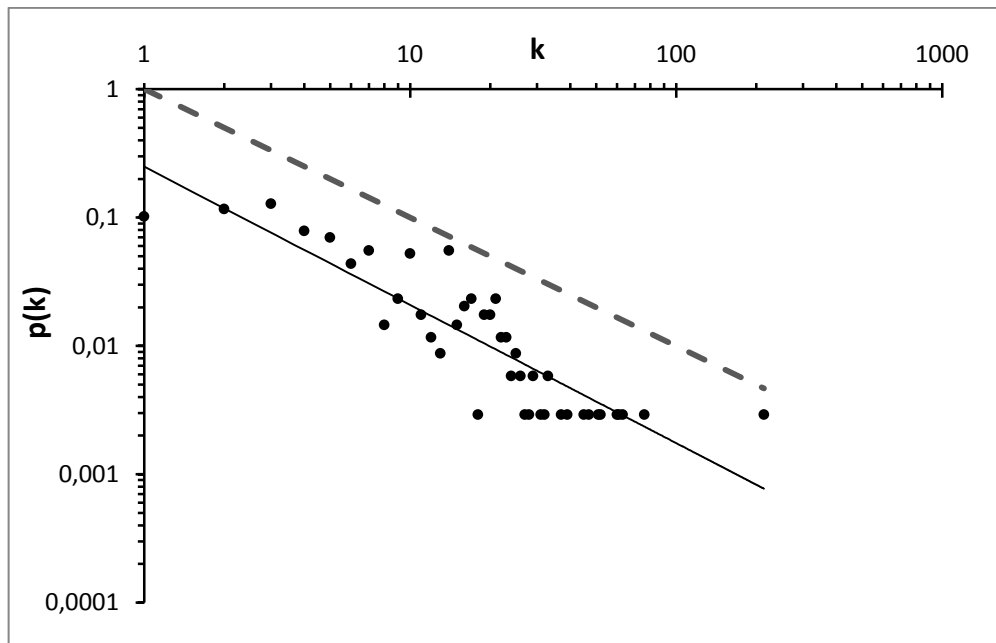
$$F(x) = 1 - (1 - bx/a)^{1/b}, \quad (4.1)$$

em que  $x$  é a variável da distribuição a ser ajustada,  $a$  é um parâmetro de escala positivo e  $b$  um parâmetro de forma [107]. De modo que a função de densidade, que denota o ajuste é dado por

$$f(x) = (1/a)(1 - bx/a)^{\frac{1-k}{k}}. \quad (4.2)$$

Uma distribuição de graus ajustada com a relação 4.2 é caracterizada por uma “cabeça” seguindo uma lei de potência e a “cauda” seguindo uma regra exponencial. Aplicando esse ajuste a distribuição de graus da Odisseia, obtivemos um ajuste melhor do que o ajuste da lei de potência (com  $\chi^2/gl = 0,01$ ). Analogamente, realizamos o mesmo ajuste para uma rede real de postagens do Facebook, e a comparamos com a distribuição da Odisseia. Os ajustes podem ser encontrados na figura 4.4, onde (A) é o ajuste Pareto Generalizada para a Odisseia, (B) o ajuste Pareto Generalizada para o Facebook e (C) o ajuste da lei de potência para o Facebook. Analisando esta imagem, mesmo a rede de postagens do Facebook, quando ajustado à lei de potência, desconsidera a “cauda” da distribuição como é o caso da Odisseia para o mesmo ajuste.

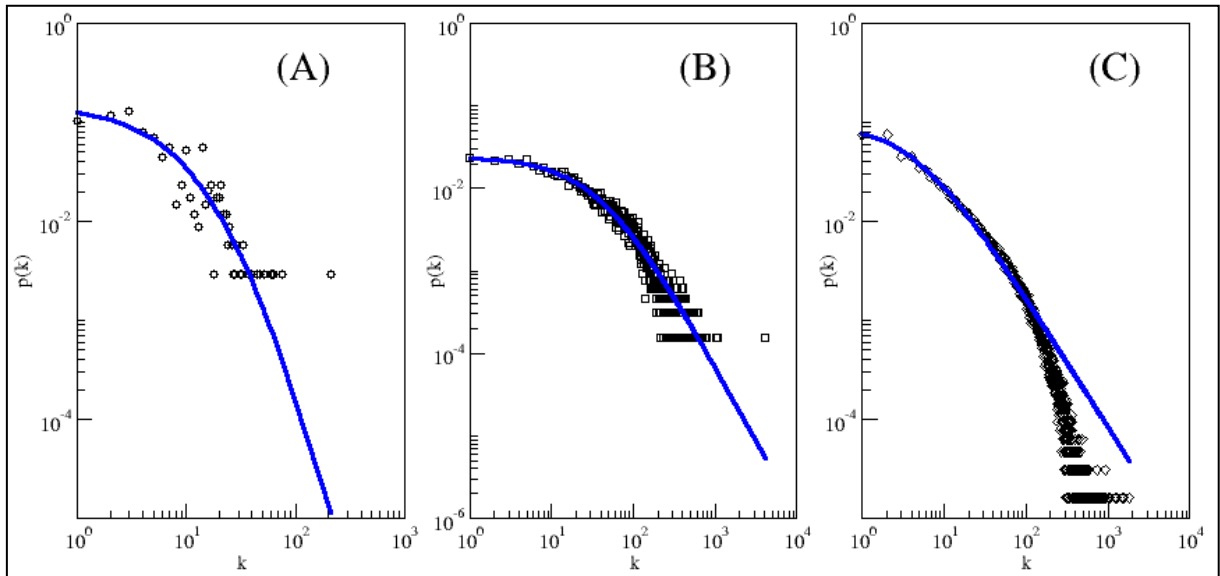
Figura 4.3 Gráfico em escala log-log que expõe a distribuição de graus como uma lei de potência. A linha cinza pontilhada é a lei de potência para expoente um e a linha preta é a regressão linear dos pontos.



Fonte: autor.

A interpretação social acerca de uma rede regida por uma distribuição de graus do tipo Pareto Generalizada, é que a rede é composta por uma parte hierarquizada, a “cabeça” da distribuição, onde as personagens menos relevantes se relacionam por meio de uma lei de potência, e uma parte não hierarquizada, que representa a relação entre as personagens mais relevantes na rede. Estes, por sua vez, coincidem com as personagens mitológicas como deuses, heróis e monstros. Uma discussão mais aprofundada sobre este resultado será realizada adiante.

Figura 4.4 Gráficos log-log de distribuições de graus (A) para Odisseia, (B) e (C) para a rede de postagens do Facebook. Onde (A) e (B) é-são o ajuste para a Pareto Generalizada e (C) o ajuste para lei de potência.



Fonte: autor.

Anteriormente foi citado que redes sociais reais tendem a ter assortatividade de grau positivas enquanto que a disassortatividade (*i. e.*, assortatividade negativa) é sinal de artificialidade na rede [32, 40, 42]. A rede social da Odisseia possui coeficiente de assortatividade de grau negativo, indicando caráter ficcional na rede. No entanto, a disassortatividade pode refletir natureza conflituosa entre as personagens. Por exemplo, o mesmo resultado foi obtido para *Ilíada* e *Táin* [36], onde os principais personagens se confrontam com inimigos que surgem apenas uma vez na estória e não têm mais nenhuma interação social, pois morrem no confronto com o protagonista. Isso resulta em personagens centrais com muitas ligações com vértices isolados que representam seus inimigos. É notável, que socialmente, estes inimigos não estão realmente isolados, mas o conto não nos oferece a informação acerca de suas relações sociais, daí a presença de disassortatividade na rede [36].

Para eliminar este problema de personagens centrais estarem conectados com muitas personagens acessórias no andamento dos eventos da estória, aqueles são removidos. Normalmente são os elementos mitológicos são tomados como deuses, heróis legendários e monstros [36]. Com isso a rede é composta apenas por pessoas que têm participação relevante na estória e ao mesmo tempo não são entidades mitológicas. A remoção destas personagens seguiu uma ordem decrescente de centralidade de intermediação, estes dados



podem ser encontrados na tabela 4.3, onde há o impacto causado na rede em termos da componente gigante e a mudança na assortatividade de grau.

É notável que a remoção das personagens estritamente mitológicas, que também são as personagens centrais, aproxima a rede ao caráter de rede social real, pois a assortatividade se torna positiva e o tamanho da componente gigante é menor que 90%. Resultados similares foram obtidos por Carron e Kenna, quando removeram as personagens mitológicas das estórias que estudaram [36]. Adversamente, é necessário analisar mutuamente quais são outras possíveis modificações causadas na rede com essa retirada de personagens, para tanto, a segunda linha da tabela 4.1 sumariza as mudanças.

Tabela 4.3 Remoção acumulada dos personagens mitológicos da Odisseia e seu impacto na assortatividade de grau e tamanho da componente gigante.

| <b>Personagem</b>             | <b>Assortatividade de grau</b> | <b>Tamanho da componente gigante</b> |
|-------------------------------|--------------------------------|--------------------------------------|
| <b>Rede completa</b>          | -0,15                          | 100%                                 |
| <b>Remoção de Odisseu</b>     | -0,07                          | 97%                                  |
| <b>+ remoção de Zeus</b>      | -0,06                          | 97%                                  |
| <b>+ remoção de Telêmaco</b>  | -0,03                          | 95%                                  |
| <b>+ remoção de Atena</b>     | -0,04                          | 93%                                  |
| <b>+ remoção de Penelope</b>  | -0,04                          | 92%                                  |
| <b>+ remoção de Menelau</b>   | 0,007                          | 92%                                  |
| <b>+ remoção de Hades</b>     | 0,03                           | 92%                                  |
| <b>+ remoção de Poseidon</b>  | 0,06                           | 91%                                  |
| <b>+ remoção de Perséfone</b> | 0,09                           | 86%                                  |

Fonte: autor.

Observando estas outras modificações, verifica-se a diminuição do número de vértices e arestas, juntamente com a diminuição do grau médio, já que os vértices mais centrais, que coincidem com os vértices mitológicos, foram removidos. A diferença entre a distância de caminho médio e a distância de caminho médio aleatório aumentaram, logo, a rede perde os aspectos de rede do tipo “small-world”. Isso implica diretamente que os elementos centrais da rede, são os atalhos descritos pelo modelo de Watt-Strogatz [18]. Mesmo com essa perda dos atalhos, a rede ainda constitui aproximadamente um fenômeno “small-world”.

Interessantemente, o diâmetro da rede, extensivamente utilizada como parâmetro fundamental em teoria de redes sociais [19, 21, 35], coincide com o diâmetro da rede da Ilíada. Vale lembrar que das redes analisadas por Carron e Kenna, Ilíada foi aquela que mais se assemelhou a redes sociais reais que, ao mesmo tempo, coincide com o diâmetro da rede da Odisseia modificada. Adicionalmente, as distribuições de grau e a distribuição dos coeficientes de aglomeração médios não tiveram modificações significativas.

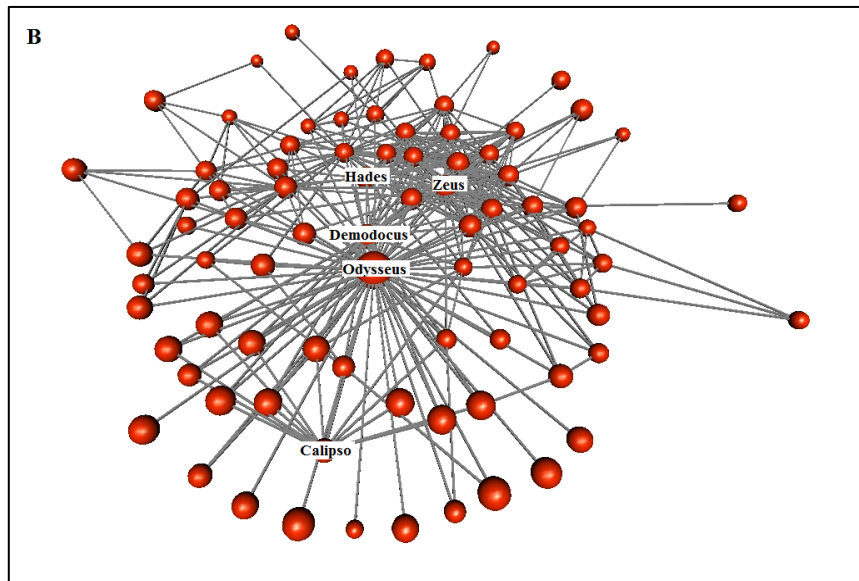
Estas manipulações e consequente interpretação levam a quase que incontornável de conclusão de que a rede pode ser concebida como uma amálgama de aspectos reais e fictícios mútuos. De modo que os aspectos fictícios são associados a personagens míticas e as personagens acessórias, em conjunto, são associadas a aspectos reais. Porquanto, de um modo geral é possível interpretar os mitos contidos na Odisseia como uma mistura de estória e história da época.

Estas características remontam o conceito do monomito discutido por Campbell em que é enfatizado que uma obra mitológica tem sua cruzada central centrada num protagonista com personagens sobrenaturais e na associação de elementos secundários aos quais são baseados nas relações sociais do contexto histórico e local [43]. Considerando estas especulações, e os resultados acerca da rede social da Odisseia, é razoável pensar que Homero poderia ter essa noção de mito e história, de modo a criar texto com conteúdos mistos, considerando estas duas naturezas. O porquê desta escolha de literatura reflete o apelo estético da literatura imbuída com elementos mitológicos o qual torna o conteúdo mais atraente às próximas gerações [48, 51].

#### **4.2.3 Análise dedutiva**

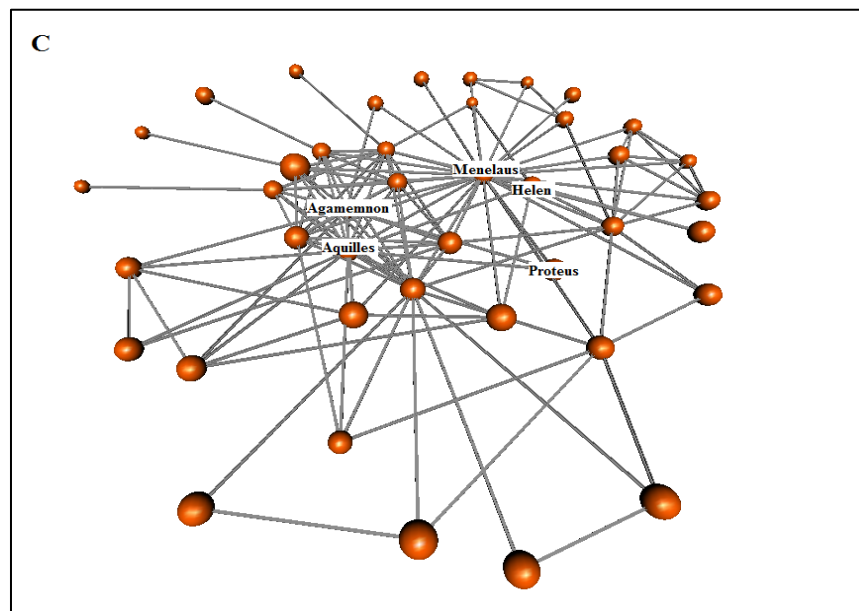
Buscando encontrar sentido dentro da rede, foram calculadas as comunidades que compõe a rede social da Odisseia. O método de procura das comunidades pelo algoritmo “walktrap” [45] permitiu identificar 32 comunidades. As imagens das duas comunidades mais importantes podem ser encontradas, respectivamente, nas Figuras 4.5 e 4.6; a primeira representa a sociedade dos deuses do Olimpo e seus descendentes e a segunda é a organização de heróis que participaram na Guerra de Troia juntamente com seus agentes relevantes.

Figura 4.5 Comunidade referente aos deuses do Olimpo e seus descendentes, assim como as entidades mitológicas contidas na Odisseia.



Fonte: autor.

Figura 4.6 Comunidade referente aos heróis que participaram na Guerra de Troia e seus agentes diretos. Também constitui a família de Menelau.



Fonte: autor.

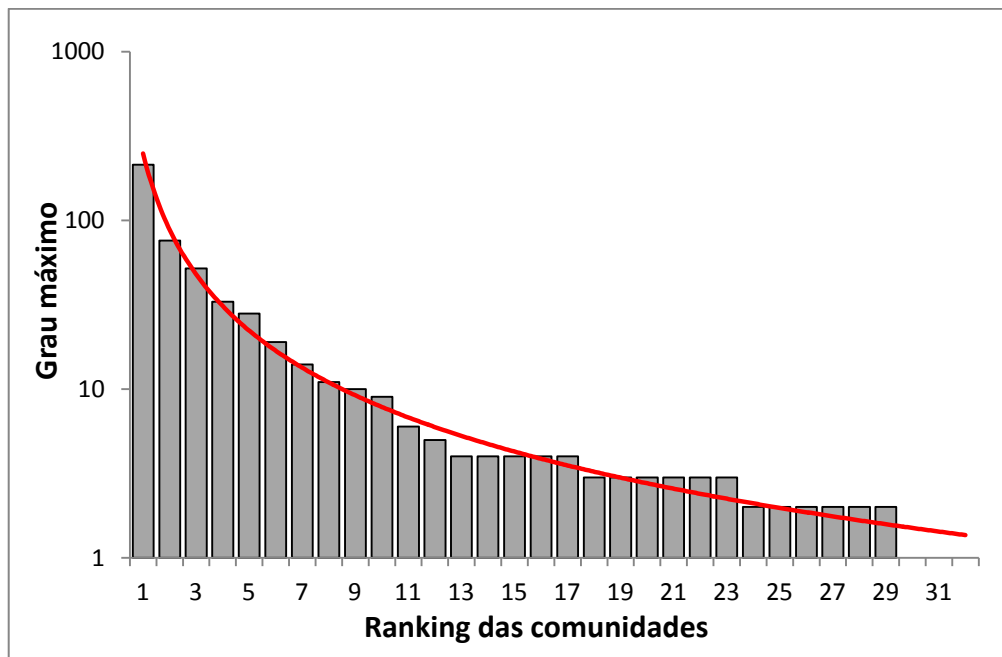
Como é de interesse saber a relação entre as comunidades, escolheu-se o grau máximo como a propriedade característica de cada comunidade, de modo a gerar o histograma ilustrado na figura 4.7. A distribuição de graus máximos de cada comunidade parece seguir uma lei de potência quando se organizam as comunidades com graus máximos decrescentes.

Isso sugere que as comunidades têm uma relação hierárquica entre elas mesmas, pois a constituição das personagens de maior grau máximo são as personagens mais centrais e assim por diante.

De algum modo, esse comportamento é esperado para uma rede de organização hierárquica tal que a hierarquização seja refletida nas comunidades subjacentes. É importante notar que a composição das comunidades reflete os eventos as quais as personagens participaram, de modo que as comunidades mais importantes albergam as personagens que participaram nos eventos mais relevantes na estória.

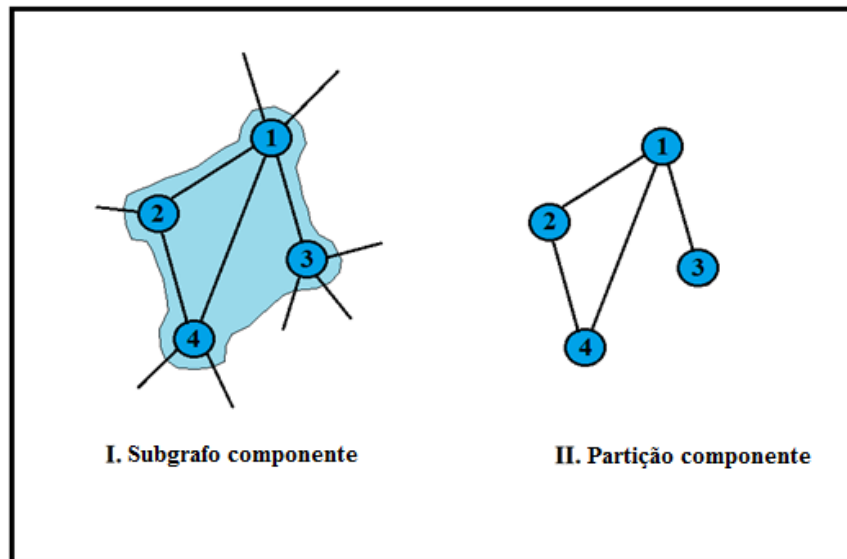
Se for considerado que as comunidades mais importantes têm  $V \geq 10$ , é possível identificar das 32, 10 comunidades mais importantes. No entanto, no cálculo das estruturas características de redes complexas, foi realizada uma diferenciação: quando foi considerada a comunidade como contínua com o resto da rede, denominou-se *subgrafo componente*, e quando se considerou a comunidade independente do resto da rede, denominou-se *partição componente*. Então, para uma mesma comunidade, há duas formas de calcular as suas propriedades estruturais. O esquema visual de cada tipo de comunidade está ilustrado na figura 4.8.

Figura 4.7 Histograma que apresenta a sequencia dos graus máximos de cada comunidade em uma ordem decrescente. A curva vermelha é a linha de tendência para a lei de potência  $1/k_{max}$ .



Fonte: autor

Figura 4.8 Ilustração de uma comunidade em suas duas acepções possíveis. I) Subgrafo componente, influenciada pela estrutura do resto da rede; II) Partição componente: em que sua estrutura é independente do resto da rede



Fonte: autor

Essa diferenciação de concepção da comunidade pode parecer trivial, mas as quantidades estruturais mudam bastante com isso. Dividindo as comunidades mais importante de A a H, temos:

- A. Contém todos os elementos dos cânticos que compõem “A Assembleia dos deuses”, (Figura 4.4);
- B. Composta pelos heróis que se excederam na Guerra de Troia e contém adicionalmente os parentes de Menelau (Figura 4.5);
- C. Sintetiza os elementos presente nos eventos que ocorrem na Jornada de volta para casa, em Ithaca. São as personagens relacionadas com os pretendentes de Penélope e seus serviçais;
- D. Comunidade composta por personagens secundárias como ninfas, monstros, e deuses menores;
- E. Composta principalmente pela família de Nestor, no episódio em que Telêmaco vai procurar por seu pai, Odisseu;
- F. Essa comunidade é uma das mais importantes na contextualização dos eventos, pois alberga as seguintes passagens de Odisseu: a Ilha dos Lotófagos, O lar de Circe, A Ilha Solar de Hiperião e a narrativa de Sila e Caribdis;

- G. Composta pelos moradores de Faecia, ilha em que Odisseu é hospedado antes de voltar para Ithaca;
- H. Composta pelos elementos da família do deus do vento Éolo, onde Odisseu e seus companheiros se hospedaram;

Os dados com os valores de cada propriedade para cada uma das 10 comunidades mais importantes nas suas duas concepções, podem ser encontradas na tabela 4.4.

Tabela 4.4 Sumário das propriedades estruturais de cada comunidade mais importante da rede da Odisseia. Tipo: S.C. (subgrafo componente), P.C. (partição componente); “*Small-world*”: se a rede for do tipo “small-world”;  $p(k)$ : padrão da distribuição de grau e *Hierárquico*: padrão de rede hierárquica; valores de assortatividade \*, representam o mesmos valores para a rede como um todo.

| Com. | Tipo | V  | $k_{max}$ | $\langle k \rangle$ | $\ell$ | $\ell_{rand}$ | C    | $C_{rand}$ | r     | “ <i>Small world</i> ” | $p(k)$          | Hierárquico |
|------|------|----|-----------|---------------------|--------|---------------|------|------------|-------|------------------------|-----------------|-------------|
| A    | S.C  | 83 | 214       | 8,30                | 2,08   | 2,41          | 0,57 | 0,23       | *     | não                    | -               | não         |
|      | P.C  | 83 | 66        | 12,8                | 2,09   | 2,09          | 0,62 | 0,27       | -0,29 | sim                    | lei de potência | não         |
| B    | S.C  | 42 | 52        | 9,10                | 2,02   | 2,09          | 0,67 | 0,26       | *     | sim                    | -               | não         |
|      | P.C  | 42 | 29        | 6,10                | 2,09   | 2,02          | 0,64 | 0,42       | -0,24 | sim                    | -               | não         |
| C    | S.C  | 73 | 76        | 13,4                | 2,05   | 2,30          | 0,61 | 0,22       | *     | não                    | -               | não         |
|      | P.C  | 73 | 46        | 11,0                | 2,12   | 2,00          | 0,59 | 0,54       | -0,28 | não                    | -               | não         |
| D    | S.C  | 10 | 10        | 3,90                | 1,86   | 2,64          | 0,66 | 0,19       | *     | não                    | -               | não         |
|      | P.C  | 10 | 6         | 2,20                | 2,00   | 1,94          | 0,28 | 0,34       | -0,40 | não                    | -               | não         |
| E    | S.C  | 11 | 19        | 9,81                | 1,30   | 1,85          | 0,65 | 0,26       | *     | não                    | -               | não         |
|      | P.C  | 11 | 9         | 6,18                | 1,30   | 1,30          | 0,70 | 0,70       | -0,23 | não                    | -               | não         |
| F    | S.C  | 25 | 28        | 9,60                | 1,78   | 2,19          | 0,66 | 0,33       | *     | não                    | -               | não         |
|      | P.C  | 25 | 19        | 6,88                | 1,82   | 1,77          | 0,66 | 0,55       | -0,26 | não                    | -               | não         |
| G    | S.C  | 20 | 33        | 14,9                | 1,31   | 2,00          | 0,74 | 0,18       | *     | não                    | -               | não         |
|      | P.C  | 20 | 16        | 12,5                | 1,31   | 1,31          | 0,82 | 0,82       | -0,12 | não                    | -               | não         |
| H    | S.C  | 13 | 14        | 14,0                | 0,92   | 1,88          | 1,00 | 0,20       | *     | não                    | -               | não         |
|      | P.C  | 13 | 12        | 12,0                | 0,92   | 0,92          | 1,00 | 1,00       | -0,01 | não                    | -               | não         |

Fonte: autor.

No processamento das comunidades por meio do algoritmo “walktrap” ao mesmo tempo em que determina o conjunto de vértices aos quais se têm uma maior densidade de arestas, também remonta as comunidades em termos dos cantos e eventos principais que as personagens compõem. Isso é possível pois a organização da rede reflete o padrão encontrada na estória. Logo, as comunidade são conjuntos de personagens próximos tanto topologicamente quanto em termos dos eventos narrados na estória.

Como é notável, o método de identificação de comunidades empregado conseguiu reunir as personagens em volta dos eventos mais importantes, que está refletida na composição da rede. Este resultado é mais um argumento em defesa acerca dos critérios de detecção da rede discutida na seção 4.1.1.

Completando a análise pela tabela 4.4, fica evidente que a comunidade A tem uma tendência especial de organização quando comparada a outras comunidades. A propriedade que aponta para isso é o alto grau médio da comunidade como uma partição componente, isso incita a imaginar que esta comunidade se comporta como uma rede social por si mesma. A característica “small-world” e livre de escala surgem, também, na comunidade A, reforçando a sua ideia de independência como rede social. No entanto, não houve padrão hierárquico nesta comunidade, de modo que a hierarquização surge apenas com o crescimento da rede no sentido de soma de comunidades, onde os elementos da comunidade A são os elementos mais centrais da hierarquização. Com isso, é possível defender que a comunidade A, é a comunidade que age como “back-bone spine” para o resto da rede, é a estrutura que estabiliza a rede mitológica da Odisseia. Outras comunidades, em si mesmas, têm pouca importância na determinação da rede como um todo. Das personagens em especial, Odisseu tem a maior centralidade de intermediação, sendo a personagem mais influente no curso dos eventos do épico. Além de Odisseu, os deuses olímpicos também têm um papel fundamental, principalmente Zeus e Atena, pois estes são os centros de decisão de muitos eventos nos quais Odisseu participa.

### **4.3 Considerações finais do capítulo**

A análise indutiva indicou que a rede social da Odisseia se aproxima do padrão “small-world”, que é organizada hierarquicamente, livre de escala e altamente aglomerada. Também foi observado que a rede é vulnerável a ataques direcionados enquanto é resiliente a ataques aleatórios. O ajuste da distribuição de graus do tipo Pareto Generalizada, permitiu uma maior cobertura dos pontos a serem ajustados, desde o começo da distribuição até o fim. O mesmo foi observado na distribuição de graus na rede de postagens do Facebook, tornando mais um argumento a favor da historicidade da Odisseia. De um modo geral, estas características são conhecidas por pertencer a redes sociais reais, apesar de dissassortatividade no grau e sua componente gigante ser maior que 90%. Embora algumas modificações foram realizadas, a assortatividade de grau e o tamanho da componente gigante puderam ser revertidas: isso foi possível com a remoção das personagens mais centrais da rede que coincide com as personagens mitológicas. Esse procedimento e resultado, indica que os elementos mitológicos, ao mesmo tempo em que dão o caráter fictício a obra, também servem

como a estrutura organizadora de toda a rede, agindo como “back-bone”, ou seja, a estrutura que centraliza outras componente da rede.

A análise dedutiva demonstrou que as comunidades, entre elas, estão organizadas de forma hierárquica, refletindo o caráter hierárquico presente entre as personagens da rede. Analogamente, a comunidade mais importante, que reflete os elementos mitológicos, é a comunidade de maior grau máximo, também é a estrutura em que foi referida como “back-bone spine”, onde a rede pode se utilizar de base.

Todos estes resultados confirmam o que se espera da análise de conteúdo das linhas poética da Odisseia: é um épico que conta histórias fabulosas por meio dos elementos humanos: sociais. Confirmando a realidade do monomito contada por Cambell [43] e corroborando com os dados analisados por Carron e Kenna [36].



## CAPÍTULO 5

### A tolerância oral como fenômeno de redes complexas

O fenômeno da tolerância oral se refere à tolerância local e sistêmica induzida no intestino delgado após a exposição do tecido mucoso a antígenos inócuos. Foi determinado recentemente que as componentes celulares e moleculares envolvidos nesse processo são de natureza complexa e de interdependência mútua, de modo que a rede de interações é concebível, mas a forma desta fenomenologia complexa ainda é desconhecida. Para tanto, este capítulo trata de um estudo aprofundado da contribuição de cada componente imunológica relacionada à tolerância oral como uma rede complexa. Também é proposta uma metodologia de cálculo de importância destas componentes baseado na importância dinâmica das mesmas, sendo utilizada a caminhada aleatória.

Este capítulo é dividido numa introdução geral sobre a imunologia básica da resposta imune adaptativa, seguindo para a descrição do modelo de caminhada aleatória, resultados analíticos e algorítmicos do modelo e suas implicações. Por fim, os resultados são comparados com a literatura que trata dos resultados experimentais de modo a gerar uma discussão acerca dos resultados por simulação.

#### 5.1 Imunologia e tolerância oral

##### 5.1.1 Conceitos gerais acerca da resposta imunológica adaptativa (AIR)

A resposta imune adaptativa, do inglês, "adaptive immune response", (AIR); constitui uma característica notável de vertebrados mandibulados, dada sua plasticidade e especificidade sistêmica. No geral, é majoritariamente composta por células secretoras de antígenos, também denominadas células B, e por um conjunto de células T que modulam distintas funções na resposta imune [53]. Por plasticidade e especificidade do AIR, refere-se à capacidade de eliminar patógenos danosos, gerar tolerância a epítopos benignos (*e. g.*, partículas de alimento), e evitar a resposta imunológica contra componentes próprios do

sistema. Estas componentes são de natureza diversa, como: componentes celulares próprios, matrizes extracelulares e partículas de alimento [54].

O timo é o órgão responsável pelo desenvolvimento e maturação de células T, mas como os peptídeos próprios e não próprios são processados por células apresentadoras de antígenos (APCs, do inglês, “antigen presenting cells”), as quais estão associadas às proteínas dos grandes complexos de histocompatibilidade (MHC, do inglês, “major histocompatibility complex”), os receptores de células T, (TCRs, do inglês, “T cells receptors”) apresentam uma região de alta variabilidade de reconhecimento de antígenos. Esta variabilidade é gerada por um rearranjo estocástico de genes relevantes. De modo que células T que reconhecem moléculas do MHC-próprias, recebem sinais de sobrevivência, e são expandidas num processo denominado seleção positiva. Por outro lado, células T que reconhecem peptídeos-próprios (moléculas externas ao MHC) passam por uma seleção negativa (morte celular) ou se tornam, naturalmente, células T regulatórias. Tais células são denominadas nTREG [55, 76]. Há ainda as células T “naïve”, aquelas que não ainda não encontraram seus antígenos cognatos, que podem sair do timo e mediar tolerância periférica a antígenos locais em tecidos mucosos como a cavidade oral e o intestino delgado. Quando expostas a antígenos, células T “naïve” se tornam células T induzidas, iTREG, que em contraste com a nTREG demonstram fenótipo transiente dependente das condições locais [57].

Dadas estas premissas, a tolerância oral se refere ao estado local e sistêmico de tolerância induzida no tecido linfóide associado ao intestino (GALT, do inglês “gut associated lymphoid tissue”) após a exposição a antígenos inócuos, como partículas de alimento [58]. As APCs presentes na lâmina própria intestinal (LP) capturam antígenos do lúmen e migram para o linfonodo mesentérico (mLN), onde iniciam a diferenciação de células T. As TREGs geradas no mLN podem retornar à LP de modo a entrar na corrente sanguínea por meio do baço onde promovem, finalmente, os efeitos sistêmicos da tolerância oral [59]. Enquanto a imunologia fez sólidos avanços em termos genéticos, moleculares e celulares das componentes envolvidas neste fenômeno [60, 61], a rede de interações entre elas ainda não foram investigadas simultaneamente. Em verdade, pouca informação se encontra disponível acerca de como esse processo ocorre, havendo apenas informações desconexas e independentes.

### **5.1.2 Tolerância oral e complexidade**

De fato, as componentes imunológicas envolvidas na tolerância oral, interagem de modo complexo em um ambiente, in vivo, único (GALT) [62]. Logo, é importante trazer à luz este problema colocando em prática os conceitos da teoria de redes complexas, de modo a proporcionar uma nova forma de analisar as interações das partes associadas ao fenômeno da tolerância oral. Com isso, é esperado gerar, heurísticamente, respostas inapreensíveis pelos meios puramente experimentais [68].

Nas próximas seções, será desenvolvido um modelo de rede complexa e dinâmica, fundamentando na descrição de quantificadores que expressem o fenômeno da tolerância oral como um fenômeno complexo sob uma dinâmica de transmissão de informações estocástica.

## **5.2 Tolerância oral e redes complexas**

### **5.2.1 Construção do modelo de rede**

Para modelar a tolerância oral, propõe-se o uso da linguagem da teoria de redes complexas para representar todas as relações entre as componentes imunológicas. Também é proposto explicar a dinâmica sistêmica como produto da dinâmica local, no contexto de rede, a dinâmica dos vértices. Com isso, uma das primeiras considerações pragmáticas é associar o fenômeno em estudo como uma resposta sistêmica a um estímulo. Tal estímulo inicial desencadeia um processo de interações especiais entre as componentes imunológicas. Seja o fenômeno em si dependente em como estas componentes estão relacionadas umas as outras e quando ocorrem. Dadas estas premissas, é proposta uma rede que capture todas as interações essenciais de modo que haja também uma quantidade associada ao estímulo em cada componente imunológico.

A rede de interações foi construída por meio da literatura disponível e adaptada para este estudo considerando apenas os elementos imunológicos associados ao GALT. De modo que componentes celulares e moleculares como linfócitos, citocinas, antígenos foram representados como vértices: os agentes da rede. As interações, relações regulatórias, transformações e estímulos foram associadas às arestas direcionadas, iniciando no agente fonte do estímulo e terminando no agente alvo do estímulo (Figura 5.1). No total, há 9 componentes moleculares e 23 elementos celulares relevantes no processo da tolerância oral associado ao

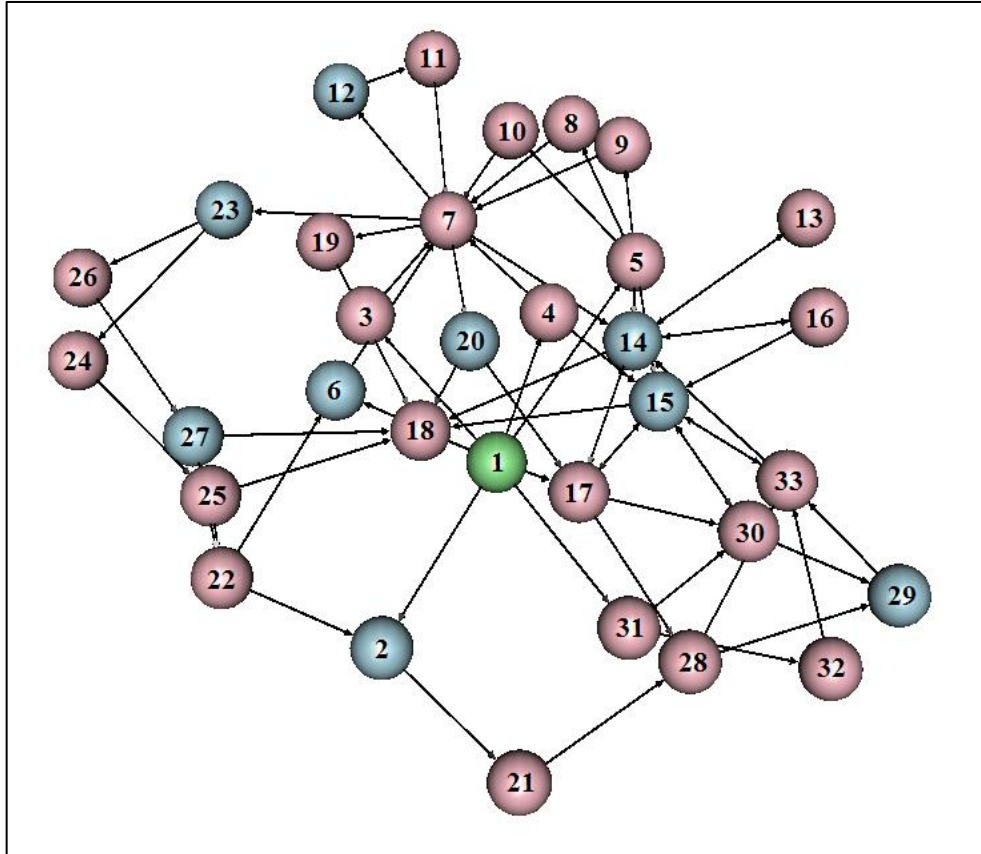
GALT, mais o elemento que inicia o estímulo na rede: o antígeno na luz intestinal. A lista de todos os elementos imunológicos está presente na Tabela 5.1. Algumas partículas externas (antígenos), como peptídeos da alimentação, podem resistir em baixo pH do fluido gástrico e as enzimas proteolíticas, alcançando o lúmen intestinal como partículas peptídicas ou mesmo proteínas intactas [63].

Qualquer antígeno em potencial pode ser complexado por IgGs e transportado para a lâmina própria intestinal através do receptor Fc neonatal; ou podem ser complexados por IgAs e transportados à lâmina pelo receptor celular, transferina (CD71) [63, 64]. Enterócitos, que são as células absorptivas que constituem a mucosa intestinal, tem um papel fundamental na captura de antígeno da luz intestinal. Pequenas moléculas (<600 Da), como proteínas desnaturadas, que podem passar entre enterócitos e se associarem a compartimentos MHC II chegando a lâmina própria na forma de exossomos [59, 65]. Analogamente, alguns antígenos podem ser transportados para Placa de Peyer por meio das células M [66].

Com isso, uma particular população de células dendríticas (DCs) residentes na lâmina própria, especificamente células CD103+, recebem e carregam a maior parte dos antígenos absorvidos pela mucosa intestinal. Estas células, então, migram para o linfonodo mesentérico onde induzem a diferenciação de células T do tipo CD4 “naïve” para células T regulatórias para os antígenos em questão. Basicamente, as células dendríticas CD103+ apresentam antígenos via o complexo de histocompatibilidade II (MHC II) combinados com o fator de crescimento e expansão  $\beta$  (TGF- $\beta$ ) e ácido retinóico (RA); estes, favorecem a diferenciação de iTREGFOXP3+ [60, 66], que em última análise são as células T regulatórias que irão atuar na tolerância destas partículas de alimentos.

A ação do RA causa uma modificação das moléculas “gut-homing” nas células T regulatórias, de modo que estas podem retornar à lâmina própria e proliferar por um mecanismo dependente de interleucinas do tipo 10 (IL-10); interleucinas são sinalizadores celulares de leucócitos. Analogamente, quando diferenciadas, iTREGFOXP3+ induzem a secreção de outro tipo de interleucina, a IL-27, por meio de células dendríticas CD11b+, que também aumentam a produção de IL-10 em células T regulatórias do tipo 1 (TR1) [67]. Ao mesmo tempo, células regulatórias e citocinas imuno-supressivas presentes na lâmina própria e no linfonodo mesentérico também promovem a tolerância sobre a inflamação no GALT em sistemas em homeostase.

Figura 5.1 Rede de relações imunológicas associada ao fenômeno da tolerância oral. Os vértices em rosa representam componentes celulares (23) e os vértices em azul as componentes humorais (9). O elemento em verde representa um antígeno genérico presente na luz intestinal. A lista dos elementos associados aos números dos vértices da rede está na tabela 5.1.



Fonte: autor

Tabela 5.1 Lista das componentes imunológicas que são os vértices da rede ilustrada na figura 5.1.

|                         |                                    |                                 |
|-------------------------|------------------------------------|---------------------------------|
| 1. Antígeno             | 12. INF-Lambda                     | 23. RA                          |
| 2. sIgG                 | 13. Th3                            | 24. $\alpha 4\beta 7$ integrina |
| 3. Células caliciformes | 14. TGF-B                          | 25. MadCAM-1                    |
| 4. Macrófagos CX3CR1    | 15. IL-10                          | 26. CCR9                        |
| 5. Enterocito           | 16. natural TregFoxP3 <sup>+</sup> | 27. CCL25                       |
| 6. sIgA                 | 17. iTregFoxP3 <sup>+</sup>        | 28. CD11c <sup>+</sup>          |
| 7. CD103 <sup>+</sup>   | 18. Naïve CD4FoxP3 <sup>-</sup>    | 29. IL-27                       |
| 8. Junções oclusivas    | 19. CCR7                           | 30. Tr1                         |
| 9. Via transcelular     | 20. IL-2                           | 31. Célula M                    |
| 10. MHC II              | 21. FcRn                           | 32. pCDs                        |
| 11. CD3 <sup>+</sup>    | 22. Célula B                       | 33. CD11b <sup>+</sup>          |

Fonte: autor

## 5.3 Caminhada aleatória como modelo de difusão de estímulos em redes complexas

### 5.3.1 Modelo dinâmico estocástico

Para entender, num sentido estatístico, a forma como se dá a interação das componentes imunológicas envolvidas na tolerância oral é proposto um modelo baseado no fluxo de caminhantes nos vértices da rede que representam a totalidade das interações. Nesta seção é proposto um estudo do comportamento assintótico da caminhada aleatória no modelo de rede complexa da tolerância oral. Sobretudo, tal processo tem pretensão de gerar quantidades que são a noção de como a tolerância oral pode ser modelada por aspectos complexos dos pontos de vista organizacional e dinâmico.

Partindo do princípio de que a rede de interações imunológicas é uma rede fixa no tempo, o número de arestas e vértices se mantém constante com o passar do tempo. Notando que a rede ilustrada na Figura 5.1 é um grafo, sem vértices desconexos, haverá sempre um caminho que passa por cada vértice ou que chegue até ele. Para representar a conectividade entre os vértices da rede, é utilizada a matriz de adjacência, de modo que se houver uma aresta que conecte o vértice  $i$  ao vértice  $j$ , o valor do elemento da matriz  $a_{ij}$  será 1, e será zero se não houver aresta.

Procurando entender o produto resultante entre a relação mútua das componentes da rede, é preciso saber o que determina tal interação mútua. Se um estímulo saiu de um vértice e atinge o outro, então deve haver uma aresta direcionada entre eles. Para tanto, é necessário redefinir o grau de saída  $k_{out}$ . Considerando um vértice genérico  $i$ , seu grau de saída pode ser denotado como

$$k_{i\ out} = \sum_j a_{ij}, \quad (5.1)$$

que constitui a soma dos elementos da linha  $i$  da matriz de adjacência. Pictoricamente, essa quantidade é expressa pelo número de arestas direcionadas que deixam o vértice  $i$ . Dados os conceitos de matriz de adjacência e grau de saída, é possível definir a caminhada aleatória para os propósitos destes estudos. Em casos gerais, a caminhada aleatória em redes complexas

pode ser entendida como um processo estocástico em tempo discreto ao qual um caminhante segue um caminho determinado pela configuração da rede complexa. Porém, diferentemente da caminhada aleatória usual, o modelo aqui proposto atribui um único vértice como fonte de criação de novos caminhantes em todos os tempos. Para formalizar esse processo, é preciso pré-determinar algumas regras de caminhadas para redes direcionadas. Estas regras podem ser organizadas na seguinte sequência:

- I. Em um tempo inicial  $t$ , um caminhante genérico  $w$  localizado inicialmente em um vértice  $i$  da rede altera sua posição para outro vértice qualquer  $j$  da rede a medida que  $t$  muda para  $t + 1$ , desde que  $i$  e  $j$  estejam conectados por uma aresta;
- II. Para os caminhantes contidos na rede, é permitido mudar de posição, de um vértice  $i$  para um vértice  $j$ , se  $j$  for um dos primeiros vizinhos de  $i$ . De modo que haja uma aresta saindo de  $i$  e alcançando  $j$ . Essa probabilidade pode ser escrita como  $h_{ij} \propto a_{ij} / \sum_j a_{ij}$ , como explorado na próxima seção;
- III. Adicionalmente, o caminhante deve respeitar a direcionalidade das arestas, seguindo da origem até a extremidade, conectando vértices;
- IV. A cada intervalo de tempo, o caminhante sempre se encontra em um vértice, nunca entre vértices.

Com isso, quando são introduzidos caminhantes na rede, cada caminhante se encontra em uma posição e obedece à dinâmica. Com isso, é conveniente escrever um vetor de posições  $W(G; t)$ . Dado um tempo inicial  $t$  para um grafo  $G$ , o tamanho do vetor  $W(G; t)$  sempre será  $N(t) = t$ , pois para cada tempo é criado um caminhante e este caminha uma vez a cada instante discreto de tempo. As componentes deste vetor recebem o rótulo do vértice em que se encontram, logo

$$W(G; t) = (w_1(t), w_2(t), w_3(t), \dots, w_{N(t)}(t)) \quad (5.2)$$

onde  $w_i(t)$  é o rótulo da posição do caminhante  $i$  inserido na rede. Com isso, a probabilidade de um caminhante mudar de  $i$  para  $j$  determina a frequência com a qual as interações imunológicas ocorrem, pois para cada tempo haverá um número de caminhantes por vértices.

À medida que o tempo do sistema passa, os caminhantes vão se concentrando nos vértices em quantidades diferentes. Para escrever o estado do sistema em termos do número de caminhantes em cada vértice, é definido um vetor de frequências  $S(G; t)$ , denotado por

$$S(G; t) = (\sigma_1(t), \sigma_2(t), \sigma_3(t) \dots \sigma_n(t)), \quad (5.3)$$

onde  $\sigma_i(t)$  representa o número de caminhantes no vértice  $i$  no tempo  $t$ . É notável que para cada instante de tempo, haverá um vetor de frequências associado, porém é necessário tornar essa quantidade mensurável entre si. Para isso, toma-se o valor relativo de caminhantes de um vértice em relação ao número de total de caminhantes na rede. A essa quantidade, atribui-se o nome de fluxo de caminhantes em  $i$ , denotado como

$$f_i(t) = \frac{\sigma_i(t)}{N(t)}. \quad (5.4)$$

Ainda, se for definido um vetor de fluxo com entradas associadas aos fluxos de caminhantes para todos os vértices da rede, tem-se um vetor que captura mutuamente o estado da rede em termos do número relativo de estímulos recebidos em cada vértice:

$$F(G; t) = (f_1(t), f_2(t), f_3(t), \dots f_n(t)). \quad (5.5)$$

De modo que altos valores de  $f_1(t)$  significam que o vértice  $i$  está sendo mais ativado num sentido imunológico. Também pode ser entendido como um maior aporte de informação passando pelo agente imunológico em termos da rede como um todo. O vetor de fluxos é de interesse estatístico para este trabalho, pois dá o estado dinâmico da rede como um todo em um instante de tempo  $t$ .

É importante notar que, quando  $t \rightarrow \infty$  o valor de fluxo  $F(G; t \rightarrow \infty)$  tende a uma distribuição estacionária se todos os vértices forem conectados. Aqui, o sentido de distribuição estacionária significa que, dado um tempo suficientemente grande para o sistema, os valores dos fluxos de cada vértice aproximam-se a um valor constante, invariante no tempo. Explorando essa propriedade dinâmica na rede, é possível criar um método de análise de fluxos que dependa de modificações estruturais.



### 5.3.2 Método de análise provido pelo modelo estocástico

O objetivo desta seção será definir um quantificador de importância de um vértice na rede de interações imunológicas. Para tanto, serão exploradas as propriedades estocásticas da matriz de transição do grafo que representa a rede de interações associada a tolerância oral. Com isso, consideremos um grafo inicial  $G$  e um tempo suficiente  $t \rightarrow \infty$ , espera-se que

$$F(G; t \rightarrow \infty) \sim \Pi(G), \quad (5.6)$$

onde  $\Pi(G)$  é a distribuição estacionária do vetor de fluxos do grafo  $G$ . Os critérios para atingir ou não a distribuição estacionária serão discutidos nas sessões seguintes, mas por hora, consideremos que o grafo seja conectado, e que para um tempo suficientemente grande, haja uma distribuição estacionária.

No entanto, for realizada alguma modificação estrutural no grafo inicial  $G$ : retirada de vértices arestas, realocação de arestas, etc. Qualquer modificação que mude a composição da matriz de adjacência. Se nomearmos  $G'$  o grafo inicial mais modificações, então é possível recalculer o vetor de fluxo para o novo grafo, de modo que haja um nova distribuição estacionária  $F(G'; t \rightarrow \infty) \sim \Pi(G')$ . Com isso, se espera que

$$\Pi(G) \neq \Pi(G'). \quad (5.7)$$

É notável que a diferença entre os vetores de fluxos num limite assintótico que tende a uma distribuição estacionária, é um quantificador estável do ponto de vista da mudança estrutural causada e o reflexo disso no processo dinâmico. Buscando definir um quantificador que sintetize em um só valor a divergência entre dois vetores de fluxos de grafos semelhantes, consideremos que a diferença mútua por componente dos vetores de fluxos:

$$\begin{aligned} f_1(t \rightarrow \infty) - f'_1(t \rightarrow \infty) &= \Delta f_i, \\ f_2(t \rightarrow \infty) - f'_2(t \rightarrow \infty) &= \Delta f_j, \\ f_3(t \rightarrow \infty) - f'_3(t \rightarrow \infty) &= \Delta f_k, \\ &\vdots \\ f_n(t \rightarrow \infty) - f'_n(t \rightarrow \infty) &= \Delta f_n. \end{aligned} \quad (5.8)$$

Os valores de  $\Delta f_i$  são a diferenças mútuas de cada componente dos vetores de fluxos entre o grafo inicial e o grafo modificado. É importante notar que a diferença de vetores de fluxos entre grafos faz sentido como quantificador quando um é fruto da modificação estrutural de outro, ou seja, deve haver modificações mínimas ou a diferença perde sentido do ponto de um quantificador dinâmico.

O estudo das possibilidades para os valores de  $\Delta f_i$ , verifica-se que podem ser negativos ou positivos dependendo em como a modificação estrutural desvie o fluxo de caminhantes em cada vértice da rede. Será negativo se a modificação causar um aumento no fluxo de caminhantes, e positivo se houve uma diminuição no fluxo de caminhantes. Como o interesse desta seção é buscar um quantificador universal para todas as diferenças, propõe-se a seguinte função que escala o valor da diferença obtida de modo que

$$p_i = \begin{cases} \frac{\Delta f_i}{f_i(t \rightarrow \infty)} & \text{para } \Delta f_i > 0 \\ \frac{\Delta f_i}{f'_i(t \rightarrow \infty)} & \text{para } \Delta f_i < 0. \end{cases} \quad (5.9)$$

onde  $p_i$  é denominado erro relativo de fluxo. Este valor escala dependendo do sinal de  $\Delta f_i$  que diz respeito da tendência de uma modificação estrutural num estado estacionário do sistema. Com os valores do erro relativo, é possível calcular o erro relativo médio

$$P(G; G'; t \rightarrow \infty) = \frac{\sum_{i=1}^n p_i}{n}. \quad (5.10)$$

O erro relativo médio,  $P(G; G'; t \rightarrow \infty)$ , é o quantificador que se buscava nesta seção. Em verdade, constitui um índice que varia de 0 a 1, onde valores próximos a 0 significam que a discrepância no fluxo de informação na rede foi pouco afetado pela modificação estrutural; e valores próximos a 1, significam que os fluxo de informação da rede foi muito alterado com a modificação estrutural. A este método é dado o nome de Método de Nocaute de Rede, que constitui a quantificação da discrepância causada num processo dinâmico via modificação estrutural.

### 5.3.3 Aplicação do método no contexto da tolerância oral

Para a rede de interações da tolerância oral, são propostos nocautes dos vértices, de modo que cada nocaute representa a ausência de um elemento celular ou humoral relacionado no fenômeno. Com isso é possível saber a importância do elemento retirado utilizando o valor do erro relativo médio entre o grafo da rede de interações “saudáveis” e o grafo sem o vértice em questão. É importante salientar que neste contexto, o vértice fonte de caminhante é o antígeno, pois é interpretado que o processo de estímulo que irá se desenvolver na rede de interação é iniciado pelo antígeno na luz intestinal. O fluxo de caminhantes, para este contexto, é o quanto um elemento na rede é ativado. Essa aproximação permite o estudo de cada componente imunológica em termos de sua parcela de importância na tolerância oral. Aqui, a tolerância oral é concebida por uma rede complexa onde ocorre um processo dinâmico de estímulos mútuos, de modo que haja um estado ideal: o vetor de fluxo da rede em seu estado completo ou saudável. A obtenção dos valores dos erros médios relativos para cada nocaute pode ser dado de suas maneiras: algoritmicamente ou explorando as propriedades algébricas do modelo estocástico.

### 5.3.4 Considerações algébricas acerca do modelo estocástico

Buscando uma forma analítica de encontrar os valores de  $f_i(t \rightarrow \infty)$  do vetor de fluxos  $F(g; t \rightarrow \infty)$  para o caso da rede de interações imunológicas da tolerância oral, é reintroduzido o conceito de matriz de transição  $T$  de um grafo. A matriz de transição é a matriz onde a entrada de seus elementos é dada pela probabilidade de um caminhante aleatório sair de um vértice  $i$  qualquer e alcançar  $j$ . Seja essa probabilidade

$$h_{ij} \propto \frac{a_{ij}}{\sum_j a_{ij}}. \quad (5.11)$$

onde o denominador é o grau de saída do vértice. Mas com isso a matriz de transição de um grafo  $G$  qualquer dada por  $T(G)$ , tem a seguinte propriedade:

$$\begin{aligned} F(G; t + 1) &= T(G)F(G; t), \\ F(G; t + 2) &= T(G)T(G)F(G; t), \end{aligned}$$

$$\begin{aligned}
& \vdots \\
& F(G; t \rightarrow \infty) \sim T(G)^\infty F(G; t), \\
& \Pi(G) \sim T(G)^\infty F(G; t).
\end{aligned} \tag{5.12}$$

As linhas da matriz  $T(G)^\infty$  representam os valores assintóticos do vetor de fluxo  $\Pi(G)$ . Essa propriedade permite obter os valores de fluxo analiticamente, independente do método iterativos descritos na próxima seção. Seja transição de estados determinados pela matriz de transição  $T(G)$  que é um processo Markoviano de  $T$  em  $G$  sob a caminhada aleatória [31].

Se for considerado o conjunto de autovalores  $\{\lambda_i\}$  da matriz de transição  $T(G)$ , de modo que a sequência de estados do vetor de fluxo para um grafo qualquer siga um processo markoviano ergódico, então sempre será possível encontrar o estado estacionário  $\Pi(G)$  no tempo assintótico. Para possibilitar estas condições, é preciso satisfazer algumas propriedades dadas pela Teoria de Matrizes não-negativas de Perron-Frobenius da matriz de transição no tempo assintótico [32]. Se a sequência dos módulos de autovalores de  $T(G)^\infty$  satisfizer a sequência  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ , de modo que  $|\lambda_1| = 1$  e  $|\lambda_i| < 1$  para todo  $2 \leq i \leq n$ , então serão satisfeitas as condições para a teoria. Sendo que o autovetor associado ao autovalor unitário será linearmente proporcional à quantidade invariante de interesse,  $\Pi(g)$ . Para calcular a matriz estocástica, consideremos que

$$\begin{aligned}
\Pi(G) &= T(G)^\infty F(G; t), \\
T(G)\Pi(G) &= \Pi(G), \\
(T(G) - I)\Pi(G) &= 0,
\end{aligned} \tag{5.13}$$

em que  $I$  representa a matriz identidade e  $\lambda = 1$  é o maior autovalor da matriz estocástica  $T(G)$ . E como a matriz  $T(G)$  é homogênea, o autovalor  $\lambda_1 = 1$  pode corresponder a mais de um vetor. Para tanto é suficiente normalizar este autovetor, resultando diretamente nos valores assintóticos do vetor de fluxo  $\Pi(G)$ . Calculando os autovalores para ao grafo correspondente a rede de interações da tolerância oral, foram encontrados  $\{|\lambda_1| = 1; |\lambda_2| = 0.958; |\lambda_3| = 0.741; |\lambda_4| = 0.718; \dots\}$ , que satisfazem a Teoria de Perron-Frobenius e possibilita a normalização dos autovetores do autovalor 1, permitindo resgatar analiticamente os valores de fluxo para um tempo  $t \rightarrow \infty$ . Os resultados analíticos estão presentes na seção 5.4 deste capítulo.

### 5.3.5 Método algorítmico de implementação do modelo

O procedimento para obtenção dos valores de fluxos e de erro relativo médio, foi um processo algorítmico implementado na Linguagem R [24]. Os detalhes do algoritmo e quantificadores podem ser encontrados no Anexo A deste manuscrito. É de interesse calcular o vetor de fluxo para um tempo grande.

No caso do método iterativo do vetor de fluxo, ou seja, método algorítmico, um segundo parâmetro é necessário: o número de repetições  $L$  das caminhadas aleatórias para um  $t$  definido. Para o método iterativo as repetições são necessárias, pois para cada processo de caminhada aleatória aplicada na rede haverá uma diferença entre os valores de fluxo de cada iteração. Como a rede é pequena, e o tempo de mistura ótima é pequeno também [32], utilizamos  $L = 10000$  e  $t = 10000$ . De modo que para cada repetição há um vetor de fluxo, logo, para cada nocaute temos:

$$\bar{F}(G; t; L) = \frac{\sum_{L=1}^L F(G; t)}{L}, \quad (5.14)$$

que é a média dos vetores de fluxos de cada repetição  $l$  do processo de caminhada aleatória da rede. Essa quantidade, denominada aqui como vetor de fluxo médio, é a medida mais adequada para se evitar os possíveis desvios de fluxos locais. A uma média de vetores de fluxos, quando a rede está completa, deu-se o nome de vetor de fluxo médio padrão. Ainda, como a rede consiste em 33 vértices, e como o vértice 1, correspondente ao antígeno na rede, é a fonte de caminhantes na rede, não pode ser nocauteado. Então é possível realizar apenas os 32 nocautes restantes, de modo a gerar 32 valores de erros relativos médios por meio dos vetores de fluxo médio de cada nocaute e o vetor de fluxo médio padrão, lembrando que cada erro relativo médio depende de um vetor de fluxo de um nocaute e do vetor de fluxo da rede saudável.

Dadas tais premissas, a comparação entre o vetor de fluxo médio padrão e qualquer outro vetor de fluxo médio produto de um nocaute, deve produzir um erro relativo médio na forma:

$$\bar{P}(G; G'; t \rightarrow 10000) \frac{\sum_{i=1}^n \bar{p}_i}{n}, \quad (5.15)$$

onde  $\bar{P}(G; G'; t \rightarrow 10000)$  é o erro relativo médio para todas as repetições  $L$ . Este é o procedimento algorítmico, isto é, iterativo de obtenção dos erros relativos que são os quantificadores de importância de cada componente na rede imunológica.

## 5.4 Resultados e discussão

### 5.4.1 Resultados

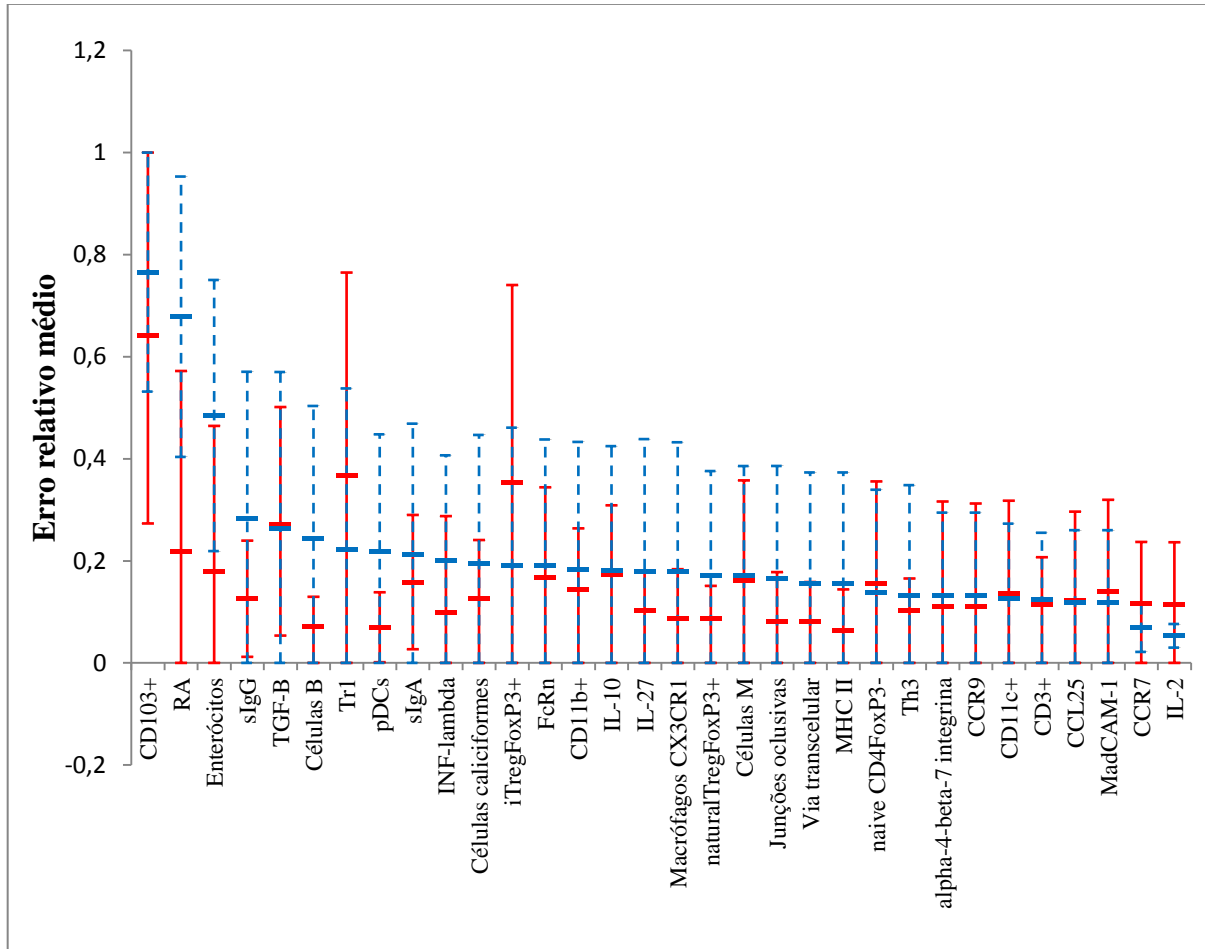
Nesta seção serão analisados simultaneamente os resultados analíticos e algorítmicos de cada erro relativo médio para cada um dos 32 nocautes, seguindo com a interpretação e implicações dos resultados na próxima seção. Para tanto, os dados dos erros relativos médios para cada caso podem ser encontrados na tabela 5.2. Os mesmos valores também podem ser encontrados na figura 5.2, juntamente com seus desvios relativos médios. É notável que o nocaute que mais causou interrupção no fluxo de caminhantes na rede foi o CD103+: as células apresentadoras de antígeno. Outros alto valores que condizem com o seu papel biológico no fenômeno são o iTREGFOXP3+, TR1, TGF- $\beta$ , RA e IL-10. Na próxima seção serão discutidas as implicações destas observações. Adicionalmente, foi calculado o erro entre os dois métodos: analítico e algorítmico, que é a média das diferenças entre os erros relativos médios. O valor encontrado foi de 7% que indica que os modelos, apesar da disparidade par a par dos mesmos nocautes que a eles se aplicaram não são consideravelmente discrepante.

Tabela 5.2 Valores dos erros relativos para os métodos analítico e algorítmico para cada nocaute da rede de interações imunológicas da tolerância oral.

| <b>Agentes imunológicos</b> | <b>Erro relativo médio<br/>(analítico)</b><br>$P(G; G'; t \rightarrow \infty)$ | <b>Erro relativo médio<br/>(algorítmico)</b><br>$\bar{P}(G; G'; t = 10^4; L = 10^4)$ | <b>Desvio</b><br>$ P - \bar{P} $ |
|-----------------------------|--|--|----------------------------------|
| <b>CD103+</b>               | 0,7  | 0,6  | 0,12                             |
| <b>Ácido retinóico (RA)</b> | 0,6  | 0,24   | 0,4                              |
| <b>Enterócito</b>           | 0,4  | 0,17   | 0,3                              |
| <b>sIgG</b>                 | 0,28   | 0,12   | 0,15                             |
| <b>TGF-B</b>                | 0,26   | 0,29   | 0,03                             |
| <b>Células B</b>            | 0,24   | 0,09   | 0,14                             |
| <b>TR1</b>                  | 0,22   | 0,3  | 0,17                             |
| <b>pDCs</b>                 | 0,21   | 0,06   | 0,14                             |
| <b>sIgA</b>                 | 0,21   | 0,15   | 0,05                             |
| <b>INF-lambda</b>           | 0,20   | 0,12   | 0,07                             |
| <b>Células calciformes</b>  | 0,19   | 0,4  | 0,21                             |
| <b>iTREGFOXP3+</b>          | 0,19   | 0,3  | 0,19                             |
| <b>FcRn</b>                 | 0,19   | 0,16   | 0,02                             |
| <b>CD11b+</b>               | 0,18   | 0,17   | 0,01                             |
| <b>IL-10</b>                | 0,18   | 0,19   | 0,01                             |
| <b>IL-27</b>                | 0,17   | 0,13   | 0,04                             |
| <b>Macrófagos CX3CR1</b>    | 0,17   | 0,08   | 0,09                             |
| <b>nTREGFOXP3+</b>          | 0,17   | 0,11   | 0,05                             |
| <b>Células M</b>            | 0,17   | 0,16   | 0,01                             |
| <b>Junções oclusivas</b>    | 0,16   | 0,08   | 0,08                             |
| <b>Vis transcelular</b>     | 0,15   | 0,08   | 0,07                             |
| <b>MHC II</b>               | 0,15   | 0,06   | 0,09                             |
| <b>naive CD4FOXP3-</b>      | 0,13   | 0,18   | 0,05                             |
| <b>TH3</b>                  | 0,13   | 0,13   | 0,00                             |
| <b>a-4-b-7 int</b>          | 0,13   | 0,07   | 0,05                             |
| <b>CCR9</b>                 | 0,13   | 0,07   | 0,05                             |
| <b>CD11c+</b>               | 0,12   | 0,16   | 0,03                             |
| <b>CD3+</b>                 | 0,12   | 0,13   | 0,01                             |
| <b>CCL25</b>                | 0,11   | 0,15   | 0,03                             |
| <b>MadCAM-1</b>             | 0,11   | 0,16   | 0,05                             |
| <b>CCR7</b>                 | 0,07   | 0,14   | 0,07                             |
| <b>IL-2</b>                 | 0,05   | 0,14   | 0,09                             |

Fonte: autor.

Figura 5.2 Representação gráfica dos erros relativos médios de cada nocaute calculados analiticamente (em azul) e algoritmicamente (em vermelho). As barras de erros são em verdade os desvios de cada nocaute.



Fonte: autor

#### 5.4.2 Discussão: síntese biológica acerca dos resultados do modelo

Nos anos 70, Niels Kaj Jerne propôs uma nova teoria para explicar a base funcional do comportamento do sistema imunológico, sugerindo, analogamente, a existência de uma rede funcional. Tal rede seria centrada em padrões de reconhecimento de idiótipos processados por linfócitos [74]. Este conceito se tornou uma faísca que iniciou o interesse de outros pesquisadores para aplicar o mesmo princípio a redes neurais como uma rede de interações funcionais. Algumas propriedades notáveis de redes imunológicas, como diversidade, autorregulação, memória, conectividade e estabilidade a perturbações; conduziram à construção de modelos matemáticos baseados em equações diferenciais, autômatos celulares e funções booleanas [75-78].

Uma das propriedades mais intrigantes do sistema imune é a autorregulação, que é o desenvolvimento de processos limitados quando o sistema é constantemente perturbado.



Neste cenário, o intestino delgado representa o local de maior contato antigênico de todo organismo, dado que alberga de 500 a 1000 espécies de micro-organismos em uma densidade de  $10^{14}$  células, correspondendo a dez vezes mais células do que as células somáticas do corpo humano e ultrapassando em duas ordens de grandeza o potencial genético humano [79]. Além da microbiota do intestino, 130 a 190 g de proteínas da alimentação são diariamente absorvidas, consistindo duplamente como maior fonte de estímulo antigênico: o GALT [80]. No entanto, quando em condições fisiológicas adequadas, a resposta prevalecente no GALT é tolerância. Tal fenômeno ocorre através de múltiplas ações das componentes imunológicas (células dendríticas, células T, plasmócitos, citocinas) interagindo em uma forma única com seu microambiente, que são compartilhados nos locais mucosos de contato antigênico [58,59, 62].

No presente trabalho, foi investigada uma rede de componentes imunes e suas respectivas interações que gerem, mantêm e regulam a tolerância oral. Na rede construída, não houve tendência de aglomeração e a distribuição de graus não segue uma lei de potência, logo não se encontrou o comportamento livre de escala. Porquanto, dado que a composição estrutural da rede não demonstrou padrões já conhecidos dos modelos de rede complexa, assumimos que a rede imunológica concebida pelos elementos do GALT, era muito parca. Ou seja, uma porção menor da rede imunológica que comporta o corpo humano inteiro. Logo, uma análise puramente estrutural é será insuficiente, daí a justificativa do uso do modelo estocástico associado ao grafo que representa a rede de interações.

No que é tocante ao modelo estocástico desenvolvido, os resultados encontrados acerca do vetor de fluxo padrão, tanto para os métodos algorítmico e analítico, sugere que as células iTREGFOXP3+ são constantemente ativadas pelo mecanismo de apresentação de antígenos, notando especialmente que os fatores TGF- $\beta$  e IL-10 são ambas citocinas supressão secretados por esta célula. As TREGs induzidas em tecidos periféricos têm papel majoritariamente supressor, ou seja, inibem células efetoras e auxiliares como TH1, TH2 e TH17, por vias moleculares. Por outro lado TREGs naturais agem preferencialmente de modo célula-a-célula [82]. Há um valor considerável de caminhantes ativando o vértice naïveCD4FOXP3- que indica que há uma dinâmica renovação de células T na lâmina própria intestinal e nos linfonodos mesentéricos. Dada a constante entrada de antígenos no sistema (proteínas alimentares, microorganismos, qualquer outro hapteno antigenóide) da luz intestinal, novas células TREG são concebidas (FOXP3-, TR1, TH3), além das células regulatórias subconjunto das células T, como células TH17 que auxilia em manter homeostase no GALT [82-84].

Adicionalmente, células TREGFOXP3<sup>+</sup> ativadas induzem secreção de IL-27 por meio da apresentação de antígeno pelas células dendríticas que estimula diferenciação e proliferação de células TR1 e produção de IL-10 [85]. Tal mecanismo age como um regulador positivo de IL-10 na lâmina própria e no linfonodo mesentérico [67]. Outros relevantes fenótipos de células dendríticas presentes no GALT também estão presentes no vetor de fluxo padrão: CD11b<sup>+</sup>, CD103<sup>-</sup>, CD11c<sup>+</sup> e C11b<sup>-</sup>. Estas células dendríticas são essenciais no processo de tolerância oral como o são na resposta imunológica de antígenos prejudiciais e patógenos. Especificamente, as células CD11b<sup>+</sup> produzem IL-10 e mantêm consistentemente uma população de células TREGs na lâmina própria, também são essenciais no processamento de antígenos na Placa de Peyer [86, 107]. Por último, células T concebidas no timo (nTREGs) são ativadas com certa relevância no vetor de fluxo padrão, permitindo a manutenção de antígenos-próprios. Analogamente, Pacholczyk *et al.* [88] demonstrou que nTREGs são cognatos a antígenos não-próprios, de modo a tolerar desta forma, antígenos comensais. Embora possam ser dispensáveis na indução de tolerância [89], a maior parte das nTREGs respondem a antígenos comensais no GALT [90].

Com isso, de uma maneira geral e qualitativa, o vetor de fluxo padrão resultante do modelo aqui proposto, representa as principais inclinações esperadas da tolerância oral, num sentido de que os resultados analítico e algorítmico para o vetor de fluxo padrão correspondem aos resultados experimentais de trabalhos que estudam cada componente separadamente. Em suma, a ativação de células TREGs, células dendríticas e produção de citocinas; é um fenômeno mútuo que pode ser estudado por meio de redes complexas.

Os resultados de nocautes dos vértices da rede de interação possibilitam conhecer a importância relativa de cada vértice no fenômeno da tolerância oral. De modo que o nocaute de células CD103<sup>+</sup> causaram grande impacto no processamento de informação na rede, correspondendo a erros relativos médio de 0,76 e 0,65 dados pelos métodos analítico e algorítmico de obtenção. Estas células dendríticas são a principal população de células dendríticas presentes na lâmina própria e linfonodo mesentérico. De relevante citação, a expressão de integrinas  $\alpha 4\beta 7$  restringem a migração de células dendríticas entre a lâmina própria e os linfonodos mesentéricos, que corresponde ao seu papel em receber antígenos da luz intestinal e encaminhar aos linfonodos onde são apresentadas a células naïveCD4<sup>+</sup> [67]. Por outro lado, as CD103<sup>+</sup> produzem citocinas de supressão como TGF- $\beta$ , IL-10 e ácido retinóico (RA) por meio da expressão de RALDH2, essencial na maturação de iTREGs. Além destas ações sinérgicas, principalmente, de TGF- $\beta$  na geração de iTREGs, o ácido retinóico

também é importante na síntese de moléculas “gut-homing” como é o caso da  $\alpha 4\beta 7$  e CCR9, permitindo TREGs retornar do linfonodo mesentérico para a lâmina própria [59, 91].

Resultados experimentais com nocautes de genes em camundongos são o análogo ao processo descrito teoricamente neste trabalho. Especialmente, camundongos nocauteados para CD103<sup>+</sup> demonstraram baixa frequência de produção de TREGs e efetores humorais de células T, demonstraram também significativa redução na expressão de moléculas “gut-homing” em células T e a tolerância oral foi grandemente afetada ou anulada [92]. Além disso, a deficiência de p38 $\alpha$  em CD103<sup>+</sup> inibe a geração de TREGs induzidas enquanto promovem o desenvolvimento de TH1 [93]. Como observados nos resultados do modelo proposto neste trabalho, a ausência de CD103<sup>+</sup> conduzem a um significativa redução na ativação de vértices que correspondem a citocinas supressivas (TGF- $\beta$ , IL-10 e IL-27), redução de todas as células TREG (FOXP3<sup>+</sup>, TR1, TH3 e TREGs derivadas do timo) e de outros subconjuntos de células dendríticas relevantes (CD11b<sup>+</sup> e CD11c<sup>+</sup>). Os valores de erro relativo associado a células caliciformes, junções oclusivas e sIgA, refletem a acumulação de antígenos por estas rotas, onde novamente, a CD103<sup>+</sup> é a responsável primária por estas capturas e carregamento destes antígenos e apresentação à células T.

Analogamente, a remoção de iTREGFOXP3<sup>+</sup> e TR1 tiveram resultados impactantes para o método algoritmo de obtenção de dados, onde o FOXP3 é um regulador chave de outras TREGs [83]. Analogamente, camundongos sem FOXP3 desenvolveram doenças autoimunes devastadoras, apresentando em sua sintomatologia: esplenomegalia, linfadenopatia, insutite, inflamação cutânea, atraso no desenvolvimento corporal e diminuição drástica na sobrevivência [94, 95]. Estes dados não discerne entre TREGs e nTREGs, no entanto evidências indicam que iTREGs são as principais responsáveis na tolerância oral [82].

A falta de marcadores específicos para células TR1 tornam difíceis de compreender sua função exata no uso clínico. Recentemente estas células foram identificadas como CD4<sup>+</sup>CD49b<sup>+</sup>LAG-3<sup>+</sup>. Onde as células TR1 são também induzidas no GALT e respondem através de altas produções de IL-10 [96]. Como no nocaute de FOXP3, ratos ausentes do gene LAG-3, responsável pela produção desta célula, exibe infiltração de leucócitos em múltiplos órgãos seguidos de doença autoimunes [97]. Há, também, um pronunciável aumento em IL-27 no modelo quando TR1 são removidas devido ao mecanismo de indução em que IL-27 secretada por células dendríticas (pDCs) que induzem diferenciação e proliferação de células TR1. No entanto, a principal citocina secretada por células TR1 é a IL-10, que pode ser dispensável na tolerância oral em baixas doses de antígenos. Também é notável que

camundongos nocauteados para TGF- $\beta$  não dá um resultado relevante na sua importância para a tolerância oral, pois vários processos metabólicos dependem desta substância, de modo que sua ausência drásticas consequências ocorrem no organismo [98, 99, 100]. No entanto, quando TGF- $\beta$  foi removido da rede de interações dado pelo modelo, houve um aumento na ativação de IL-10 e então diferenciação de células TR1 que pode agir como contra-mecanismo para manter a tolerância no GALT. Também, os níveis de iTREGs foram um pouco menores e os níveis de TH3 são completamente anulados, pois o fator TGF- $\beta$  é um elemento essencial no desenvolvimento destas células T [101-103].

## **5.5 Considerações finais do capítulo**

O modelo baseado em uma rede complexa de interações imunológicas possibilitou o uso de dinâmica estocástica para gerar quantificadores e descrever o fenômeno da tolerância oral assim como o comportamento de suas componentes. A maior limitação enfrentada pelo modelo e seu poder de previsão é a ausência do fator de peso nas interações imunológicas. No entanto, a construção partindo de dados qualitativos permitiu corroborar os significados dos resultados providos pelo modelo e os resultados experimentais. O presente modelo aliado a outros modelos disponíveis na literatura formam um ponto inicial de referência para a construção de modelos mais acurados, baseado em dados quantitativos [104, 105].

## CAPÍTULO 6

### Conclusões e trabalhos futuros

O conceito abstrato de emergência denota simplesmente a necessidade de modelar certos sistemas reais por meio de grafos. Com o uso da teoria de grafos e dinâmica estocástica foi possível analisar sistemas reais em dois estudos de casos muito discrepantes em substância, mas semelhantes na forma com que suas partes se relacionam. De modo que, estudando sucintamente a teoria de grafos (Capítulo 2), foi possível identificar as ferramentas mais importantes na descrição e análise de redes complexas reais, como é o caso da rede social da Odisseia. Em segundo lugar, o estudo de processos markovianos possibilitou encontrar alguns resultados formais de interesse para a caminhada aleatória em redes complexas. Para este trabalho, a quantidade invariante associada ao vetor de fluxo de caminhantes (*i. e.*, informação) de um grafo sob regime estocástico (Capítulo 3).

O estudo de caso de uma rede mitológica, da Odisseia de Homero; permitiu a aplicação dos conceitos de teoria de redes complexas de modo a gerar identificadores de padrões de redes sociais (Capítulo 4). Deste estudo foi possível verificar que a rede mitológica da Odisseia pode ser interpretada por aspectos reais e fictícios em termos de redes sociais, pois a rede possui quantificadores que indicam estruturas reais, como: “small-world”, livre de escala, vulnerabilidade a ataques direcionados e resiliência a ataques aleatórios. Em contrapartida a outros quantificadores que sinalizaram que a rede também tem caráter fictício, como: não ser hierárquica, possuir componente gigante maior que 90% dos vértices e assortatividade de grau negativa. Isso torna a rede ambígua, pois há tanto indicadores de redes reais quanto de redes fictícias. Para tentar contornar esse paradoxo, foram retirados da rede, personagens reconhecidamente mitológicos, o que demonstrou seu caráter real, como: a presença de hierarquidade, componente gigante menor que 90% dos vértices e assortatividade de grau positiva. Com isso, é possível dizer que os personagens fictícios, em conjunto, tornam a rede mitológica da Odisseia uma rede mais semelhante a uma rede fictícia, ao passo que a rede resultante da retirada destas personagens, se aproxima de uma rede real.

Em segundo lugar, a aplicação da dinâmica estocástica em rede complexa como modelo de estudo para o fenômeno da tolerância oral, permitiu a descrição de um novo

quantificador baseado no fluxo de informação na rede (Capítulo 5). Onde uma quantidade invariante num tempo suficientemente grande foi possível de ser obtida computacionalmente e analiticamente. Dado que, se a matriz de transição da rede observada satisfizer as condições da teoria de Perron-Frobenius, a rede deve atingir um estado estacionário num tempo transiente. De modo que este estado está associado a um vetor de fluxo estacionário, calculado pela equação de autovalor da matriz de transição inicial. Simultâneo a este formalismo, foi implementado um algoritmo de caminhada aleatória que descreve todo o processo até a obtenção do vetor de fluxo que represente o estado estacionário para um tempo suficientemente grande computacionalmente ( $t = 10^4$ ). Em suma, os resultados teóricos aqui encontrados e discutidos, corroboraram os resultados experimentais, de modo que as componentes imunológicas mais importantes no modelo, correspondem também às componentes mais importantes do ponto de vista biológico-experimental, a saber: ITREGFOXP3+, TGF-B, RA, Tr1 e CD103+. De modo que a metodologia de análise de redes imunológicas desenvolvidas neste estudo, também permite a previsão de resultados experimentais não realizados, dado que a construção da rede é baseada na revisão bibliográfica de trabalhos experimentais.

Realizados estes dois estudos, fica claro que o modelo de grafo é poderoso quando os estados gerais do sistema depende fundamentalmente da forma com que suas partes se relacionam. Isso justifica a aplicação e a interdisciplinaridade que a teoria de redes complexas possibilita em estudos de sistemas reais complexos, mas de naturezas distintas.

Como trabalhos futuros, é pretendido um aprofundamento na aplicação da teoria de categorias como modelos mais gerais que possam ser transmitidos ao conceito de redes complexas. Para o aprofundamento de estudos de redes sociais extraídos de textos, pretende-se a comparação da rede da Odisseia com uma rede de postagens do Facebook, de modo a salientar suas semelhanças e diferenças. Para o estudo de redes imunológicas, pretende-se o aumento da rede em termos de vértices, ou seja, agentes imunológicos que determinam a fenomenologia da resposta imune adaptativa. Também é pretendido encontrar o peso nas arestas direcionadas tornando o modelo mais próximo do real. De uma forma geral, os estudos contidos nesta dissertação são um início para o aprofundamento da modelagem de fenômenos diversos por meio de redes complexas e dinâmica estocástica.

## Referências Bibliográficas

1. DESCARTES, R. *Discours de la méthode*. Tradução de Maria Ermantina Galvão. São Paulo: Martins Fontes, 1996.
2. LIVINGTON, D. M. *The Master of Light: A Biography of Albert A. Michelson*. New York: Scribner, 1973.
3. GLEICK, J. *Chaos: Making a new science*. New York: Penguin, 1988.
4. SCHRODINGER, E. *What is life?* New York: Cambridge University Press, 1944.
5. PÉREZ, R. G. *Introducción a la metafísica (Aristóteles y Santo Tomás de Aquino)*. Madrid: Rialp, Imp, 1978.
6. KIM, J. Making Sense of Emergence. *Philosophical Studies*, v. 95, p. 3-36, 1999.
7. WALDROP, M. M. *Complexity : the emerging science and the edge of order and chaos*. London: Viking, 1992.
8. HILBERT, M.; LOPEZ, P. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, v. 332, p. 60–65, 2011.
9. VAN MIEGHEM, P. *Graph Spectra for Complex Networks*. New York: Cambridge University Press, p. 13, 2011.
10. VAN STEEN, M. *Graph Theory and Complex Network*. Amsterdã: Publicado pelo autor, p. 18, 2010.
11. RUOHONEN, K. *Graph Theory*. Disponível em <math.tut.fi/~ruohonen/english.html>. Acesso em 19 de fevereiro de 2014.
12. ALBERT, R.; BARABÁSI, A-L. *Statistical mechanics of complex networks*. *Rev. Mod. Phys.*, 74, p. 47-97, 2002.
13. NEWMAN, M.; BARABÁSI, A-L.; WATTS, D. J.; *The structure and dynamics of networks*. Princeton University Press: Princeton and Oxford, 2006.
14. ERDOS, P.; RÉNYI, A. *On random graphs*. *Publ. Math.* 6, 290, 1959.
15. ERDOS, P.; RÉNYI, A. *On the evolution of random graphs*. *Publ. Math. Inst. Hung. Acad. Sci.* 5, p. 17, 1960.
16. ERDOS, P.; RÉNYI, A. *On the strength of connectedness of a random graph*. *Bull. Inst. Int. Stat.* 38, p. 343, 1961.
17. BOLLOBÁS, B. *Random Graphs*. *Lon. Math. Soc. Mono.: Acad. Press: London*. p. 447, 1985.
18. WATTS, D. J.; STROGATZ, S. H. *Collective dynamics of 'small-world' networks*.

- London: Nature 393, p. 440, 1998.
19. ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. *Diameter of the World-Wide Web*. London: Nature 401, p. 130. 1999.
  20. BARABÁSI, AL.; ALBERT, R. *Emergence os scaling in random networks*. London: Science 286, p. 30, 1999.
  21. WASSERMAN, F.; FAUST, K. *Social Network Analysis*. Cambridge: University Press, Cambridge, 1994.
  22. RAVASZ, E.; BARABÁSI, A-L. *Hierarchical organization on complex networks*. Phys. Rev. E. 67, p.35, 2003.
  23. FREEMAN, L. C. *A set of Measures of Centrality Based on Betweenness*. Sociometry, 40, p. 35, 1977.
  24. R DEVELOPMENT CORE TEAM. *R: A Language and Environment for statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, p.409, 2009.
  25. PONS, P.; LATAPY, M. *Computing communities in large networks using random walks*. Phys. Soc. arXiv:physics/0512106v1, 2004.
  26. HUGUES, B. D. *Random Walks on Random Enviroments: Volume 1*. Oxford: University Press, 1995.
  27. BACHELIER, L. *Theorie de La Especulation*. Ann. Sci. Ecole. Norm. Sup. 27, p. 21-86, 1900.
  28. EINSTEIN, A. *Brownian motion*. Ann. d. Physik. XXI, p. 371, 1905.
  29. WEISSTEIN, E. W. *Random Walk-1-Dimensional*. Tech. Rep.: Wolfram Web resource, 2003.
  30. TOMÉ, T.; OLIVEIRA, M. J. *Dinâmica Estocástica e Irreversibilidade*. São Paulo, Edusp, 2001.
  31. SANTOS, R. J. *Cadeia de Markov*. Imprensa Universitária, Universidade Federal de Minas Gerais, 2006.
  32. SENETA, E. *Non-negative matrices and Markov chains (2<sup>nd</sup> ed.)*. New York: Springer Verlag, 1981.
  33. NEWMAN, M. E. J. *Networks, An Introduction*. New York: Oxford University Press, 2010.
  34. AMARAL, L. A.; SCALA, A.; BARTHÉLÉMY, M.; STANLEY, H. E. *Classes of small-world networks*. Proc. Natl. Acad. Sci. U.S.A. 97, 2000.
  35. ONNELA, J.-P. *Taxonomies of networks from community structure*. arXiv:5731.
  36. CARRON, P.M.; KENNA, R. *Universal properties of mythological networks*. E.P.L.



- 99, 2012.
37. NEWMAN, M. E. J. *Assortative mixing in networks*. Phys. Rev. Lett. 89, 2002.
  38. NEWMAN, M. E. J.; PARK, J. *Why social networks are different from other types of networks*. Phys. Rev. E. 68, 2003.
  39. DOROGOVTSEV, S. N.; GOLTSEV, A. V.; MENDES, J. F. F. *Pseudofractal scale-free web*. Phys. Rev. E. 65, 2002.
  40. GLEISER, P. M. *How to become a superhero*. J. Stat. Mech., 2007.
  41. ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. *Error and attack tolerance of complex networks*. Nature 406, 2000.
  42. ALBERICH, R.; MIRO-JULIA, J.; ROSELLÓ, F. *Marvel Universe looks almost like a real social network*. Cond-mat, 2002.
  43. CAMPBELL, J. *The Hero with a Thousand Faces*. Princeton: Princeton University Press, 1949.
  44. MENDES, M.; RODRIGUES, A. M. *Odisseia de Homero*. São Paulo: Editora da Universidade de São Paulo, 2000.
  45. PONS, P.; LATAPY, M. *Computing communities in large networks using random walks*. Phys. Soc., 2005.
  46. FREEMAN, L. C. *A set of Measures of Centrality Based on Betweenness*. Sociometry 40, p. 35, 1977,
  47. RIEU, E. V. *The Odyssey*. London: Penguin Classics, 1946.
  48. RIEU, E. V. *Introduction to The Odyssey*. London: Penguin Classics, 2003.
  49. KORFMAN, M. “Was There a Trojan War?” Archeology 57, p. 36, 2004.
  50. KRAFT, J. C.; RAPP, G.; KAYAN, I.; LUCE, J. V. *Harbor areas at ancient Troy: Sedimentology and geomorphology complement Homer’s Iliad*. Geology, 31, p. 163-166, 2003.
  51. PALMEIRA, E. D.; CORREIA, M. A. *Odisseia de Homero*. Portugal: Livraria Sá da Costa Editora, 1994.
  52. ELSON, D.K.; DAMES N.; MCKEOWN, K.R.. *Extracting Social Networks from Literary Fiction*. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, p. 138–147, Uppsala, Sweden, 2010.
  53. MUPHY, K. M.; TRAVERS, P.; WALPORT, M. *Janeway’s Immunobiology* 8<sup>th</sup> ed. New York: Taylor & Francis, 2012,

54. BURT, T. D. *Fetal Regulatory T cells and Peripheral Immune Tolerance In Utero: Implications for Development and Disease*. Am. J. Repro. Immunol. 69, p; 346-358, 2013.
55. KOSMRLJ, A.; CHAKRABORTY, A. K.; KARDAR, M., SHAKNOVICH, E. I. *Thymic Selection of T cell Receptors as an Extreme Value Problem*. Physical Review Letters 103, 2009.
56. GORDON, J.; MANLEY, N. R. *Mechanism of Thymus organogenesis and morphogenesis*. Development 138, p. 3865-3878, 2011.
57. HUEHN, J.; POLANSKY, J. K.; HAMANN, A. *Epigenetic control of FOXP3 expression: the key to a stable regulatory T cell lineage?* Nat. Ver. Immunol. 9, p; 83-89, 2009.
58. WEINER, H. L.; DA CUNHA, A. P.; QUINTANA, F.; WU, H. *Oral Tolerance*. Immunol. Rev. 241, p. 241-259, 2011.
59. PABST, O.; MOWAT, A. M. *Oral tolerance to food protein*. Mucosal Immunology., p. 1-8, 2012.
60. OHKURA, N.; KITAGAWA, Y.; SAKAGUSHI, S. *Development and maintenance of Regulatory T cells*. Immunity 38, p. 414-423, 2013.
61. ÁLVAREZ, B. S.; RANEROS, A. B.; ORTEGA, F.; LARREA, C. L. *Epigenetic modulation of the immune function: A potential target for tolerance*. Epigenetics 8, p. 694-702, 2013.
62. PEARSON, C.; UHLIG, H. H.; POWRIE, F. *Lymphoid microenvironments and innate lymphoid cells in the gut*. Trends in Immunol. 33, p. 289-296, 2012.
63. MÉNARD, S.; BENSUSSAN, N. C.; HEYMAN, M. *Multiple facets of intestinal permeability and epithelial handling of dietary antigens*. Mucosal Immunology, p. 1-13, 2010.
64. SUTHERLAND, D. B.; FAGARASAN, S. *IgA synthesis: a form of functional immune adaptation extending beyond gut*. Current Op Immunol. 24, p. 261-268, 2012.
65. MIRON, N.; CRISTEA, V. *Enterocytes: active cells in tolerance to food and microbial antigens in the gut*. Clin Exp Immunol. 167, p. 405-412, 2011.
66. NAGATANI, K.; KOMAGATA, Y.; ASAKO, K.; TAKAYAMA, M., YAMAMOTO, K. *Antigen-specific regulatory T cells are detected in Peyer's patches after the interaction between T cells and dendritic cells loaded with orally administered antigen*. Immunobiol. 216, p. 416-422, 2011.

67. JAENSSON, E.; URONEN-HANSON, H.; PABST, O.; EKSTEEN, B.; TIAN, J.; COOMBES, J. L.; BERG, P. L.; DAVIDSSON, T.; POWRIE, F.; JOHANSSON-LINDBOM, B.; AGACE, W. W. *Small intestinal CD103 + dendritic cells display unique functional properties that are conserved between mice and humans*. *The J. Exp. Med.* 205, p. 2139-2149, 2008.
68. AWASTHI, A.; CARRIER, Y.; PERON, J. P. S.; BETTELLI, E.; KAMANAKA, M.; FLAVELL, R. A.; KYCHROO, V. K.; OUKKA, M.; WEINER, H. L. *A dominant function for interleukin 27 in generating interleukin 10-producing anti-inflammatory T cells*. *Nat. Immunol.* 8, p. 1380-1389, 2007.
69. BAR-YAM, Y. *Dynamics of complex systems*. Massachusetts: Addison-Wesley, 1997.
70. VARELA, F. G.; MATURANA, H. R.; URIBE, R. B. *Autopoiesis: the organization of living systems; its characterization and a model*. *Biosystems* 5, p. 185-205, 1974.
71. ANDOLFI, G.; FOUSTERI, G.; ROSSETI, M.; MAGNANI, C. F.; JOFRA, T. *Enforced IL-10 expression confers type 1 regulatory T cell (Tr1) phenotype and function to human CD4<sup>+</sup> T cells*. *Mol. Ther. Nucleic. Acids.* 20, p. 1778-1790, 2012.
72. LI, M.; SANJABI, S.; FLAVELL, R. A. *Transforming growth factor- $\beta$  controls development, homeostasis, and tolerance of T cells by regulatory T cell-dependent and independent mechanisms*. *Immunity* 25, p. 455-471, 2005.
73. HALL, J. A.; GRAINGER, J. R.; SPENCER, S. P.; BELKAIDE, Y. *The Role of Retinoic Acid in Tolerance and Immunity*. *Immunity* 35, p. 13-22. 2011.
74. SÖDERQVIST, T. *The troubled life of Niels Jerne*. Yale: Yale University Press, 2003.
75. CELADA, F.; SEIDEN, P. E. *A computer model of cellular interaction in the immune system*. *Immunol. today* 13, p. 56-62, 1992.
76. DE BOER, R. J.; PERELSON, A. S. *Size and connectivity as emergent properties of a developing immune network*. *J. Theor. Biol.* 149, p. 381-424, 1991.
77. RIBEIRO, L. C.; DICKMAN, R.; BERNARDES, A. T.; VAZ, N. M. *Dynamic stability in random and scale-free B lymphocyte networks*. *Phys. Rev. E.* 75, p. 1-26, 2007.
78. THAKAR, J.; PATHAK, A.K.; MURPHY, L.; ALBERT, R.; CATTADORI, I. M. *Network model of immune responses reveals key effectors to single and co-infection dynamics by a respiratory bacterium and a gastrointestinal helminth*. *Plos Comput. biol.* 8, p. 1-19, 2012.
79. SOMMER, F.; BÄCKHED, F. *The gut microbiota – masters of host development and physiology*. *Nat. Rev. Microbio.*, p. 1-12, 2003.

80. BRANDTZAEG, P. *Development and basic mechanisms of human gut immunity.* Nutr. Rev. 56, p. 5–18, 1998.
81. BARABÁSI, A. L.; OLTVAI, Z. N. *Network Biology: Understanding the Cell's Functional Organization.* Nat. Revi. Genet. 5, p. 101-113, 2004.
82. BILATE, A. M.; LAFAILLE, J. J. *Induced CD4<sup>+</sup>FoxP3<sup>+</sup> regulatory T cell in Immune Tolerance.* Annu. Rev. Immunol. 30, p. 733-58, 2012.
83. HIROTA, K.; TURNER, J. E.; VILLA, M.; DUARTE, J. H.; DEMENGEOT, J. *Plasticity of Th17 cells in Peyer's patches is responsible for the induction of T cell-dependent IgA responses.* Nat. Immunol. 14, p. 372-380, 2013.
84. DENNING, T.L.; WANG, Y. C.; PATEL, S. R.; WILLIAMS, I. R.; PULENDRAN, B. *Lamina propria macrophages and dendritic cells differentially induce regulatory and interleukin 17-producing T cell responses.* Nat. Immunol. 8, p. 1086-1094, 2007.
85. POT, C.; APETOH, L.; AWASHI, A.; KUCHROO, V. K. *Molecular Pathways in the Induction of Interleukin-27-Driven Regulatory Type 1 Cells.* J. Interf. & Cyto. Resear. 30, p. 381-388, 2010.
86. WELTY, N. E.; STALEY, C.; GHILARDI, N.; SADOWSKY, M. J.; IGYÁRTÓ, B. Z. *Intestinal lamina propria dendritic cells maintain T cell homeostasis but do not affect commensalism.* J. Exp. Med., p. 2011-2024, 2013.
87. NAGATANI, K.; KOMAGATA, Y.; ASAKO, K.; TAKAYAMA, M.; YAMAMOTO, K. *Antigen-specific regulatory T cells are detected in Peyer's patches after the interaction between T cells and dendritic cells loaded with orally administered antigen.* Immunobiol. 216, p. 416-422, 2011.
88. PACHOLCZYK, R.; KERN, J.; SINGH, N.; IWASHIMA, M.; KRAJ, P.; IGNATOWICZ, L. *Nonsel-self-antigens are the cognate specificities of FoxP3<sup>+</sup> regulatory T cells.* Immunity 27, p. 493-504, 2007.
89. MUCIDA, D.; KUTCHUKHIDZE, N.; ERAZO, A.; RUSSO, M.; LAFAILLE, J. J. *Oral tolerance in the absence of naturally occurring Tregs.* J. Clin. Invest. 115, p. 1923-1933, 2005.
90. CEBULA, A.; SEWARYN, M.; REMPALA, G. A.; PABLA, S. S.; MCINDOE, R. A. *Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota.* Nature 497, p. 258-265, 2013.
91. SCOTT, C. L.; AUMENUNIER, A. M.; MOWAT, A. *Intestinal CD103<sup>+</sup> dendritic cells: master regulators of tolerance?* Trends in Immunol. 32, p. 412-419, 2011.

92. SEMMRICH, M.; PLATINGA, M.; SVENSSON-FREJ, M.; URONEN-HANSSON, H.; GUSTAFSSON, T. *Directed antigen targeting in vivo identifies a role for CD103<sup>+</sup> dendritic cells in both tolerogenic and immunogenic T-cell responses.* Mucosal Immun. 5, p. 150-160, 2012.
93. HUANG, G.; WANG, Y.; CHI, H. *Control of T cell fates and immune tolerance by p38 $\alpha$  signalling in mucosal CD103<sup>+</sup> dendritic cells.* J. Immunol. 15, p. 650-659, 2003.
94. LAHL, K.; LODDENKEMPER, C.; DROUIN, C.; FREYER, J.; ARNASON, J. *Selective depletion of FoxP3<sup>+</sup> regulatory T cells induces a scurfy-like disease.* J. Exp. Med. 204, p. 57-63, 2007.
95. ZHANG, B.; SUN, C.; QU, Y.; ZHANG, A.; LIU, J. *Deficiency of mouse CD4<sup>+</sup>CD25<sup>+</sup>Foxp3<sup>+</sup> regulatory T cells in xenogenic pig thymus-grafted nude mice suffering from autoimmune diseases.* Cell & Molec. Immunol. 5, p. 325-332, 2008.
96. GAGLIANI, N.; MAGNANI, C. F.; HUBER, S.; GIANOLINI, M. E.; PALA, M. *Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells.* Nat. Med. 38, p. 739-746, 2013.
97. WOO, S. R.; TURNIS, M. E.; GOLDBERG, M.V.; BANKOTI, J.; SELBY, M. *Immune inhibitory molecules LAG-3 and PD-1 synergistically regulate T cell function to promote tumoral immune escape.* Cancer Res. 72, p. 917-927, 2012.
98. GONNELLA, P. A.; WALDNER, H. P.; KODALI, D.; WEINER, H. L. *Induction of low dose oral tolerance in IL-10 deficient mice with experimental autoimmune encephalomyelitis.* J. of Autoimmunity 23, p. 193-200, 2004.
99. SHEVACH, E. M.; DAVIDSON, T. S.; HUTER, E. N.; DIPAOLO, R. A.; ANDERSSON, J. *Role of TGF- $\beta$  in the induction of FoxP3 expression and T regulatory Cell function.* J. Clin. Immunol. 28, p. 640-646, 2008.
100. BÖTTINGER, E. P.; LETTERIO, J. J.; ROBERTS, A. B. *Biology of TGF- $\beta$  in knockout and transgenic mouse models.* Kidney international 51, p. 1355-1360, 1997.
101. REYNOLDS, L. A.; MAIZELS, R. M. *In the absence of TGF- $\beta$  signaling in T cells, fewer CD103<sup>+</sup> regulatory T cells develop, but exuberant INF- $\gamma$  production renders mice more susceptible to helminth infection.* J. immunol. 189, p. 1113-1117, 2012.
102. CHANG, S. Y.; CHA, H. R.; CHANG, J. H.; KO, H. J.; YANG, H. *Lack of retinoic acid leads to increased langerin-expressing dendritic cells in gut-associated lymphoid tissues.* Gastroenterology 138, p. 1468-1478, 2010.
103. BUETTNER, M.; PABST, R.; BODE, U. *Lymph node stromal cells strongly influence immune response suppression.* Eur. J. Immunol. 41, p. 624-633, 2011.

104. THAKAR, J.; PILIONE, M.; KIRIMANJESWARA, G.; HARVILL, E. T. ALBERT, R. *Modeling systems-level regulation of host immune responses*. Plos Comput. Biol. 3, p. 1022-1039, 2007.
105. THAKAR, J.; PATHAK, A. K.; MURPHY, L.; ALBERT, R.; CATTADORI, I. M. *Network model of immune responses reveals key effectors to single and co-infection dynamics by a respiratory bacterium and a gastrointestinal helminth*. Plos Comput. boil. 8, p. 1-19, 2012.
106. HOSKING, J. R. M.; WALLIS, J. R. *Parameter and quantile estimation for the generalized pareto distribution*. Technometrics 3, vol. 29, p. 339-349, 1987.
107. CHOULAKIAN, V.; STEPHENS, M. A. *Goodness-of-fit tests for the generalized pareto distribution*. Technometrics 4, vol. 43, p. 478-484, 2001.

## ANEXO A

# Linhas de comando da Linguagem interpretativa R: algoritmo de caminhada aleatória e análise de fluxos de informação

### B1. Script para a caminhada aleatória da rede de interações imunológicas

# K quais knockouts # L número de amostragens # T tempo de caminhada

```
library(igraph, Matrix)
```

```
imuno<-read.table(file="m.imuno.dat")
```

```
g<-graph.data.frame(imuno, directed=TRUE)
```

```
K<-33
```

```
nomes<-V(g)$name
```

```
matriz<-as.matrix(V(g)$name)
```

```
freqs<-as.matrix(V(g)$name)
```

```
for(k in 1:K){KO<-nomes[k]; gko<-delete.vertices(g, c(KO));
```

```
  matriz<-cbind(matriz, c(V(gko)$name));{
```

```
L<-1000;
```

```

for(l in 1:L){W<-1; T<-1000; B<-c(); A<-vector(length=W, mode="numeric");

    imuno<-read.table(file="m.imuno.dat");g<-graph.data.frame(imuno,
directed=TRUE);

for(i in 1:W){A[i]<-1}

for(t in 1:T){W<-length(A);{

    for(i in 1:W){

        if(degree(gko, mode=c("out"), A[i])==0){A[i]<-A[i];}

        else{ vizinhos<-neighbors(gko, A[i], mode="out");

            if(length(vizinhos)==1){A[i]<-vizinhos[1];}

                else{A[i]<-sample(vizinhos,1)}}}

}; A<-c(A,1)}; if(l==1){B<-c(B, A); write.table(B, file="B.dat")}

    else{B<-c(read.table(file="B.dat")); B<-c(B$x, A);

        write.table(B, file="B.dat"); B<-c(); A<-c()};

N<-33;

C<-vector(length=N, mode="numeric");

B<-c(read.table(file="B.dat"));

B<-B$x;

```



```
for(n in 1:N){N[n]<-n ; S<-length(B);

for(i in 1:S){

    if(B[i]==N[n]){C[n]<-C[n]+1}

    else(C[n]<-C[n]+0)

    }};

C[1]<-0;

f<-C/sum(C); freqs<-cbind(freqs, c(f))}

freqs<-freqs[,-1]

matriz<-matriz[-33,-1]

write.table(freqs, file="freqsKO.dat", sep=" ", dec=",")

write.table(matriz, file="nomesKO.dat", sep=" ", dec=",")
```