

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Identificação de propagadores influentes em redes complexas

André Luiz Barbieri

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

André Luiz Barbieri

Identificação de propagadores influentes em redes complexas

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
Janeiro de 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

B236i Barbieri, André Luiz
Identificação de propagadores influentes em redes
complexas / André Luiz Barbieri; orientador
Francisco Aparecido Rodrigues. -- São Carlos, 2022.
90 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2022.

1. Redes complexas. 2. Epidemias. 3. Rumores. 4.
Propagadores. 5. Acessibilidade. I. Rodrigues,
Francisco Aparecido, orient. II. Título.

André Luiz Barbieri

Identification of influential spreaders in complex networks

Thesis submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Francisco Aparecido Rodrigues

USP – São Carlos
January 2022

*Dedico este trabalho aos meus pais, esposa e minha filha
que, desde o início, foram minha força e coragem para seguir.
Em especial, dedicado ao meu amigo professor e aos cientistas
do Instituto de Ciências Matemáticas e de Computação (I C M C).*

AGRADECIMENTOS

Os agradecimentos mais importantes são dirigidos à Don Dieguito, Filipinho, Cabelão, Pedro, Ana, Adriele, Kakaia, Nato, Guilherme e todos aqueles que contribuíram para que a produção deste trabalho acadêmico fosse possível.

Agradecimento especial é direcionado ao grande amigo Chico que sempre me apoiou.

*“Alguma coisa incrível, em algum lugar,
está esperando para ser descoberta.”
(C. Sagan)*

RESUMO

BARBIERI, A. L. **Identificação de propagadores influentes em redes complexas**. 2022. 90 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Esta tese apresenta um estudo de dinâmicas de propagação em redes reais e artificiais. A teoria das redes tem sido muito utilizada para se modelar sistemas reais. Identificar os propagadores mais importantes nas redes é vital para entender e controlar as capacidades de disseminação do sistema. Trabalhos mais recentes têm sugerido que identificar propagadores influentes é dependente da dinâmica a ser aplicada. Os mais importantes propagadores não são necessariamente os mesmos quando se trata de analisar contágio social e propagação de rumor. Mostramos que diferentes medidas (grau, *coreness*, etc.) podem identificar diferentes vértices influentes. Neste trabalho, analisamos como medidas de centralidade se correlacionam com as capacidades de propagar doenças e rumores em diferentes redes sintéticas e do mundo real (ambas espaciais e não-espaciais). Propõe-se generalizar a medida de acessibilidade por caminhada aleatória como uma nova medida de centralidade e derivar expressões analíticas desta última medida para simples configurações de rede. Nossos resultados mostram que, para redes não-espaciais, as centralidades *k-core* e grau são mais correlacionadas com eventos como epidemia, enquanto o grau médio da vizinhança e as medidas de proximidade e acessibilidade são mais relacionadas à dinâmica de rumor. E para as redes espaciais, verificou-se que a medida de acessibilidade se destaca entre as medidas de centralidade para quase todos os casos, não dependendo da dinâmica considerada. Uma consequência desta análise é que estudos realizados anteriormente em redes aleatórias sintéticas não podem ser generalizados para redes espaciais.

Palavras-chave: Redes complexas, Epidemias, Rumores, Propagadores, Acessibilidade.

ABSTRACT

BARBIERI, A. L. **Identification of influential spreaders in complex networks.** 2022. 90 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

This thesis presents a study of propagation dynamics in real and artificial networks. Network theory has been widely used to model real systems. Identifying the most important propagators in the networks is vital to understanding and controlling the dissemination capabilities of system. More recent work has suggested that identifying influential propagators is dependent on the dynamics to be applied. The most important propagators are not necessarily the same if it comes to analyzing rumor propagation and social contagion. We show that different measures (*coreness*, degree, etc.) can identify different important vertices. In this work, we analyze how measures of centrality correlate with the capacities to propagate diseases and rumors in different real-world and synthetic networks (both non spatial and spatial). It is proposed to generalize the accessibility measure of random walk as a new centrality metric and to derive the analytical expressions of this last metric for simple configurations of network. Our analyze show that, for non spatial networks, the centralities degree and *k-core* are more correlated with events such as epidemics, while the average of degree of the neighborhood and measures of proximity and accessibility are more related to the dynamics of rumor. And for spatial networks, it was found that the measure of accessibility stands out among the measures of centrality for almost all cases, not depending on the dynamics considered. A result of this analysis show that studies carried out previously on synthetic aleatory networks can not be generalized to the spatial-networks.

Keywords: Complex networks, Epidemics, Rumors, Spreaders, Accessibility.

LISTA DE ILUSTRAÇÕES

Figura 1	– (a) Rede de amizades de um clube de karate (ZACHARY, 1977). (b) Histograma normalizado da distribuição de grau.	30
Figura 2	– Exemplos de coeficiente de clustering calculado para o nó em destaque. As linhas tracejadas representam a falta de conexão entre vizinhos.	31
Figura 3	– Rede de rodovias pelo Brasil. As rodovias representadas pelas arestas e as cidades representadas pelos pontos.	32
Figura 4	– $\langle k_{nn} \rangle$ da Internet entre 97 e 99.	34
Figura 5	– (a) Rede aleatória <i>ER</i> . (b) Distribuição de grau de rede de mil nós, usando probabilidade $p = 0,2$. Média de várias simulações.	37
Figura 6	– Exemplos de redes variando p no modelo mundo pequeno com grau médio $\langle k \rangle = 4$	38
Figura 7	– (a) Gráfico da medida de coeficiente de clustering cc e do comprimento médio ℓ usando diferentes limites de probabilidade p aplicados ao modelo <i>WS</i> . (b) Distribuição de grau para um grafo com $N, 1000, \langle k \rangle, 200$, usando $p = 0,01$	38
Figura 8	– (a) Rede do tipo livre de escala criada através do modelo <i>SF</i> . (b) Gráfico com escala logarítmica para distribuição de grau (em azul) e da distribuição cumulativa (em vermelho) para grafo de mil nós.	40
Figura 9	– Diagrama de fluxo do modelo de epidemia <i>SIR</i>	44
Figura 10	– Evolução do modelo <i>SIR</i>	45
Figura 11	– Rede Santa74 do Facebook: (a) distribuições de probabilidades do percentual máximo de nós infectados no modelo <i>SIR</i> , (b) distribuições de índice central M_i e (c) distribuição de índice central M_i considerando-se apenas o cluster que contém os propagadores mais influentes e do cluster com o restante vértices.	53
Figura 12	– Visualização da rede Santa74 do Facebook. Os nós na curva em vermelho pertencem ao cluster com os maiores valores de M_i . Esta visualização baseia-se num paradigma “força-dirigida” por Kawai e Kamada (KAMADA; KAWAI, 1989), disponível para Cytoscape (SHANNON <i>et al.</i> , 2003).	55
Figura 13	– Porcentagem máxima de indivíduos contaminados após procedimentos de imunização com base no grau, <i>k-core</i> , cluster mais central e aleatório para a rede de email (a) - (c), rede Political blogs (d) - (f), rede Smith60 do Facebook (g) - (i) e rede Santa74 do Facebook (j) - (l). Diferentes valores de taxa de difusão (β) e recuperação (μ) foram considerados.	56

Figura 14 – Avaliação da propagação da doença na rede árvore artificial. (a) de visualização de rede: os nós em forma de vermelho a fragmentação mais central. Esta visualização baseia-se no paradigma de “força-dirigida” conforme implementado por Kamada e Kawai (KAMADA; KAWAI, 1989; SHANNON <i>et al.</i> , 2003). Em (b) e (c): porcentagem máxima de indivíduos infectados após os procedimentos de imunização baseada em hubs, cluster mais central e escolha aleatória de nós para $\mu = 0.3$ e $\mu = 0.6$, respectivamente.	57
Figura 15 – Gráfico de conceito da medida acessibilidade (valores baseados na Equação 5.4) na rede do clube de Caratê, Zachary (ZACHARY, 1977). Nós centrais apresentam a maior acessibilidade. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	61
Figura 16 – Acessibilidade calculada para a estrela (a partir das Eqs. 5.12 e 5.11), grafos completamente conectados, linha e anel em comparação com o valor máximo obtido nos grafos (ER), redes (BA), e redes livre de escala espaciais (S S F). N é o número de nós. Cada ponto é uma média de +50 redes de $\langle k \rangle \approx 4$. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	63
Figura 17 – Relação entre propagação e conectividade, k , acessibilidade, α e k-core, k_c , tomando todos valores dos parâmetros ($\beta \times \mu$ para a difusão de epidemia (SIR) e $\lambda \times \delta$ para a difusão de rumor (MT TP)), em redes de rodovias da Alemanha e dos EUA. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	69
Figura 18 – Relação entre propagação e conectividade, k , acessibilidade, α e k-core, k_c , tomando todos os valores dos parâmetros ($\beta \times \mu$ para a difusão de epidemia (SIR) e $\lambda \times \delta$ para a difusão de rumor (MT TP)), para as redes sociais : e-mails e blogs de políticos. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	70
Figura 19 – A porcentagem dos agentes recuperados no modelo de propagação SIR ($\mu = 1.0, \beta = 0.3$) conforme as medidas locais das rede de rodovia dos EUA: (a) conectividade; (b) coeficiente de clustering; (c) grau médio dos vizinhos; (d) <i>betweenness</i> ; (e) Page Rank; (f) k-shell; (g) autovetores e (h) proximidade. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	72
Figura 20 – A porcentagem de <i>stiflers</i> no modelo de rumor MT ($\delta = 1.0, \lambda = 0.3$) conforme as medidas locais para a rede de rodovias dos EUA: (a) a conectividade; (b) coeficiente de clustering; (c) grau médio dos vizinhos; (d) <i>betweenness</i> ; (e) Page Rank; (f) k-shell; (g) autovetores e (h) proximidade. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	72
Figura 21 – A porcentagem dos agentes recuperados no modelo de propagação SIR ($\mu = 1.0, \beta = 0.3$) conforme a medida de acessibilidade para redes rodoviárias do (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	73

Figura 22 – A porcentagem de <i>stiflers</i> no modelo de rumor MT ($\delta = 1.0, \lambda = 0.3$) conforme medida de acessibilidade para redes rodoviárias (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	74
Figura 23 – Grafo das redes de roddovias do (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Pontos em vermelho tem maior valor de acessibilidade. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	75
Figura 24 – Distribuição de probabilidade da medida de acessibilidade nas redes de rodovias de Japão, Alemanha, Estados Unidos e Inglaterra. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	75
Figura 25 – Distribuição de probabilidade da medida de acessibilidade nas redes sociais Advogato, email, blogs políticos e Google+. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	76
Figura 26 – Variação da medida de acessibilidade e o percentual de agentes recuperados no modelo de propagação SIR ($\beta = 0.3, \mu = 1.0$) para os grafos (a) Advogato, (b) de políticos, (c) mensagens, (d) +Google. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	77
Figura 27 – Variação da medida de acessibilidade e o percentual de stiflers no modelo de rumores MT ($\delta = 1.0, \lambda = 0.3$) para os grafos (a) Advogato, (b) de políticos, (c) mensagens, (d) +Google. Figura extraída de (ARRUDA <i>et al.</i> , 2014).	78
Figura 28 – Fraçao de infectados ao longo do tempo usando diferentes valores das medidas τ_R e h_{thr} em redes mundo pequeno.	83

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Subrotina para obter curva do modelo de epidemia SIR	44
---	----

LISTA DE TABELAS

Tabela 1 – Exemplos de redes e suas Assortatividades.	34
Tabela 2 – Propriedades de rede sem escala: número de nós (N), grau médio ($\langle k \rangle$), média do coeficiente de clustering ($\langle cc \rangle$), coeficiente de assortatividade (x) e médias dos caminhos mais curtos ($\langle l \rangle$).	52
Tabela 3 – Medidas topológicas das redes da base de dados.	68
Tabela 4 – Relação entre centralidade e o percentual de agentes recuperados (SIR) ou o percentual de <i>stiflers</i> (MT). As medidas são conectividade (k), coeficiente de clustering (cc), <i>betweenness</i> (B), média do grau dos vizinhos (r), Page Rank (π), autovetores (x), k-shell (k_c), proximidade (C) e acessibilidade (α). Em negrito, os maiores valores.	79
Tabela 5 – Relaçãp entre centralidade e o percentual de stiflers (MT) ou o percentual de recuperados (SIR). As medidas são conectividade (k), coeficiente de clustering (cc), <i>betweenness</i> (B), média do grau dos vizinhos (r), Page Rank (π), autovetores (x), k-shell (k_c), proximidade (C) e acessibilidade (α). Em negrito, os maiores valores.	80

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos	26
1.2	Contribuições	26
1.3	Descrição de capítulos	26
2	MEDIDAS E MODELOS DE GERAÇÃO DE REDES	29
2.1	Medidas	29
2.1.1	<i>Grau</i>	29
2.1.2	<i>Distribuição de grau</i>	30
2.1.3	<i>Coefficiente de clustering</i>	30
2.1.4	<i>Distância</i>	31
2.1.5	<i>Grau de proximidade closeness centrality</i>	33
2.1.6	<i>Betweenness centrality</i>	33
2.1.7	<i>Assortatividade</i>	33
2.1.8	<i>Grau médio de vizinhos</i>	35
2.1.9	<i>Page Rank</i>	35
2.1.10	<i>Centralidade por autovetor</i>	36
2.1.11	<i>Coreness</i>	36
2.2	Modelos de redes complexas	36
2.2.1	<i>Redes aleatórias</i>	37
2.2.2	<i>Modelo Mundo Pequeno</i>	38
2.2.3	<i>Redes livre de escala</i>	39
2.2.4	<i>Redes espaciais</i>	41
3	PROPAGAÇÃO DE EPIDEMIAS E RUMORES	43
3.1	Modelo <i>SIR</i>	43
3.1.1	<i>Equações do modelo SIR</i>	45
3.2	Modelo <i>S I S</i>	46
3.3	Modelo de rumor	46
4	IDENTIFICAÇÃO DE PROPAGADORES INFLUENTES ATRAVÉS DE MINERAÇÃO DE DADOS	49
4.1	Propagação da epidemia	50
4.2	Maximização da esperança	50

4.3	Conjunto de dados	51
4.4	Resultados e discussão	52
4.5	Conclusões	58
5	IDENTIFICAÇÃO DE PROPAGADORES IMPORTANTES USANDO MEDIDAS DE CENTRALIDADE	59
5.1	Acessibilidade de caminhada aleatória generalizada	60
5.1.1	<i>Acessibilidade de grafo em estrela</i>	<i>61</i>
5.1.1.1	<i>Decompondo matriz</i>	<i>62</i>
5.1.2	<i>Acessibilidade de grafo em anel</i>	<i>64</i>
5.1.3	<i>Acessibilidade em grafo completo</i>	<i>65</i>
5.2	Propagação de epidemias e rumores	67
5.3	Base de dados	67
5.4	Análise dos parâmetros	68
5.5	Redes espaciais	71
5.5.1	<i>Redes de rodovias</i>	<i>73</i>
5.6	Rede não espacial	74
5.7	Conclusões	76
6	CONCLUSÃO	81
	REFERÊNCIAS	85

INTRODUÇÃO

Nossa sociedade é formada por pessoas que interagem, formando uma grande rede complexa (BARABÁSI, 2003). Nessa rede, são propagadas epidemias, rumores e diversos tipos de informações (PASTOR-SATORRAS *et al.*, 2015). Essa propagação está fortemente relacionada com a estrutura da rede, de modo que propriedades na organização, como assortatividade ou distância entre pessoas, pode afetar a propagação. Assim sendo, compreender como a organização da rede pode influenciar na propagação é ainda um problema fundamental, de modo que alterando-se a estrutura, pode-se controlar processos epidêmicos. Um exemplo típico ocorreu recentemente, quando quarentenas foram necessárias para controlar a propagação do novo coronavírus (BABAEI *et al.*, 2021).

Um outro aspecto fundamental, é a identificação dos principais propagadores, que são vértices com alto poder de disseminação de uma informação. Esses vértices afetam a propagação, interferindo de forma fundamental tanto na propagação de epidemias, como de rumores (KITSAK *et al.*, 2010). Por exemplo, trabalhos recentes mostraram que podem haver super propagadores do novo coronavírus, acelerando a disseminação da doença (SNEPPEN; SIMONSEN, 2020). Ademais, em redes sociais online a identificação de propagadores influentes é importante para controlar a disseminação de notícias falsas e controlar a desinformação nessas redes. No entanto, uma questão fundamental é como identificar tais propagadores.

Nesse trabalho, estamos interessados em desenvolver métodos para identificação dos principais propagadores de rumores e doenças em redes complexas. Vamos analisar como a estrutura da rede afeta a propagação e determinar as medidas de centralidade que permitem identificar os principais propagadores. De fato, esse é um trabalho importante que tem recebido atenção da comunidade recentemente (ver (LÜ *et al.*, 2016) e (RADICCHI; CASTELLANO, 2016)). Comparamos diferentes medidas para identificar os principais propagadores e propomos uma nova medida de centralidade, denominada acessibilidade de caminha aleatória. Ademais, usamos métodos de mineração de dados para associar os principais propagadores a outliers, em

termos da estrutura da rede. Aplicamos nossa metodologia em redes reais a artificiais.

1.1 Objetivos

O principal objetivo deste trabalho é introduzir métodos para identificar os propagadores mais influentes em redes complexas. Para alcançar esse objetivo, as seguintes atividades foram implementadas:

1. Implementar medidas para caracterização de redes complexas.
2. Implementar modelos clássicos de geração de redes complexas.
3. Implementar as dinâmicas de difusão de epidemias e rumores sobre redes.
4. Utilizar métodos de mineração em dados para classificar vértices em redes a fim de identificar os principais propagadores.
5. Comparar medidas de centralidade para identificação dos principais propagadores.
6. Propor uma nova medida de centralidade que seja mais precisa na identificação dos principais propagadores.

1.2 Contribuições

Quanto às contribuições principais presentes neste trabalho, citamos:

- Desenvolvimento de um método para identificação dos principais propagadores baseando-se em métodos de clustering de dados e medidas de centralidade.
- Introdução da medida acessibilidade generalizada em redes, que permite a identificação dos propagadores mais influentes.

1.3 Descrição de capítulos

Iniciamos o estudo sobre redes complexas no Capítulo 2, onde será visto os conceitos fundamentais sobre redes complexas. Será feita uma revisão dos tipos de grafos, formas de representá-los e como obter medidas básicas que servem para descrever relações de conexões dos vértices. Veremos os modelos que desenvolvidos para criação de redes complexas, pelos quais podemos estudar inúmeros sistemas complexos.

Nos seguintes 2 capítulos, foram analisadas redes artificiais e reais com a intenção de quantificar a centralidade de vértices e relacionar com resultados obtidos das dinâmicas de propagação. Já no Capítulo 4 analisamos redes sociais como: a rede de emails trocados entre

usuários numa universidade; a rede do Facebook de usuários de algumas universidades em um determinado período; a rede de blogs políticos; e a rede artificial do tipo árvore baseada nas medidas estruturais das demais redes. O capítulo 5 mostra a análise de redes espaciais reais e artificiais, além de incluir as redes sociais, sendo elas: redes de rodovias, extraídas com nosso algoritmo, o qual é capaz de obter as redes através de mapas rodoviários como do Japão, Inglaterra, Estados Unidos e Alemanha; redes espaciais artificiais, baseadas no modelo livre de escala espacial e de Waxman; redes sociais, como a Advogado, Google Plus; e a rede artificial baseada no modelo de Barabási e Albert. Analisamos computacionalmente a dinâmica de propagação sobre todas estas redes. Várias métricas de centralidade foram obtidas destas redes, com o intuito de classificar os vértices mais centrais. Modelos de propagação de epidemia e rumor foram simulados, a fim obtermos a eficiência de propagação de cada vértice. Foram relacionadas às medidas de centralidade com a capacidade de propagar dos vértices, assim foi possível identificar os vértices mais influentes em relação à dinâmica de propagação empregada. Verificamos a existência de uma forte correlação entre a medida de acessibilidade e a eficiência de propagação do vértice, permitindo usar esta nova métrica para descrever a influência de vértices em redes.

Para finalizar, mostraremos as conclusões e também os possíveis trabalhos futuros.

MEDIDAS E MODELOS DE GERAÇÃO DE REDES

Nesta primeira seção do capítulo as medidas mais comuns serão apresentadas para caracterizar a organização de redes complexas. Na Seção 2.2 serão descritos os modelos mais utilizados para a geração das redes.

2.1 Medidas

Especialmente em redes, as métricas são importantes, logo que, a maior parte das redes complexas são compostas por 10^3 (ou mesmo, 10^6) nós, e somente visualizar não seria bastante para obter informação sobre a organização da rede. Veremos as medidas mais conhecidas e que melhor descrevem a importância de nós dentro da rede.

2.1.1 Grau

Dentre as medidas de redes, a mais comum é conhecida como grau, ou conectividade, e é muito simples de ser obtida. Quando não existirem múltiplas conexões entre os nós, o grau e a conectividade são similares, e apenas este caso será tratado nesta tese. A conectividade de um nó i é dada pela quantidade de arestas que chegam a ele (para de redes que não são dirigidas), se denota por k_i , e pode ser obtido pela equação seguinte,

$$k_i = \sum_{j=1}^N a_{ij}. \quad (2.1)$$

onde a_{ij} é membro (i,j) da matriz A de adjacência.

Como o grau é uma métrica localizada, para descrever a organização de redes podemos usar uma métrica global que pode ser obtida pela média do quantidade de conexões entre os nós,

e que se chama grau médio (conectividade média) $\langle k \rangle$,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (2.2)$$

2.1.2 Distribuição de grau

A estrutura das conexões de uma rede é descrita pela distribuição das arestas. Esta distribuição é denotada por $P(k)$, k é o grau, e $P(k = d)$ designa a probabilidade de ser encontrado um nó com grau igual a d conexões.

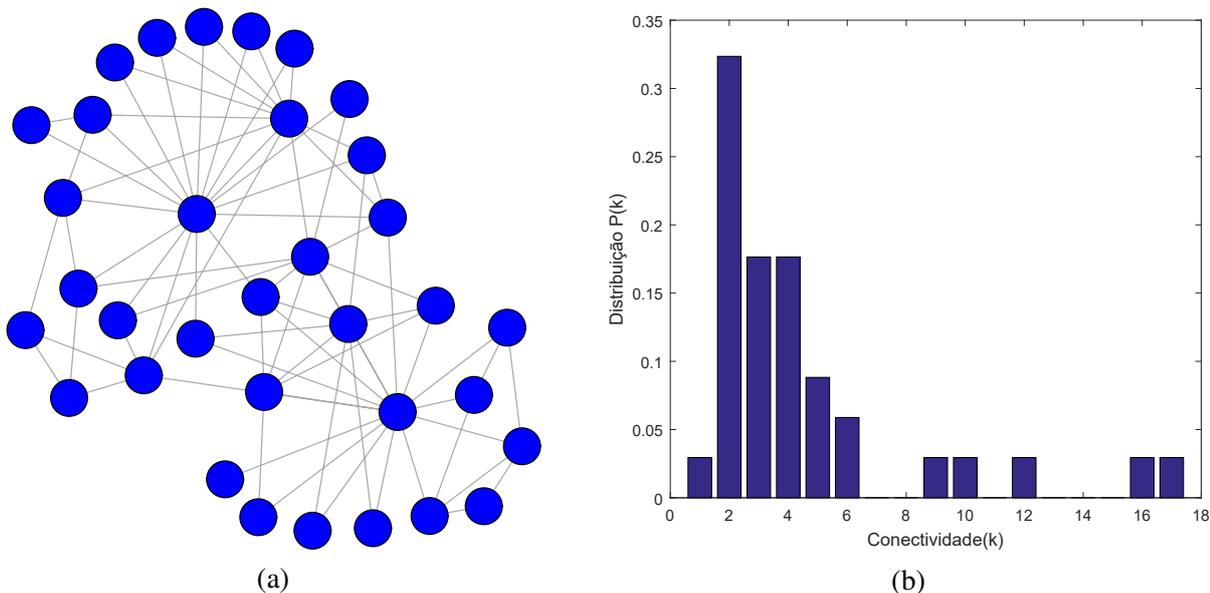


Figura 1 – (a) Rede de amizades de um clube de karate (ZACHARY, 1977). (b) Histograma normalizado da distribuição de grau.

Fonte: Elaborada pelo autor.

O gráfico da distribuição de grau é basicamente um histograma normalizado do número de nós observando todos os graus k , como na Figura 1(b), e para isso se utilizou-se a rede de amizades de um clube de karate (ZACHARY, 1977) como a Figura 1(a).

2.1.3 Coeficiente de clustering

Redes normalmente tem ciclos de ordem três (três vértices conectados) como podemos ver na rede da *Caenorhabditis elegans* e também na rede de colaboração de atores (WATTS; STROGATZ, 1998). Uma forma de identificar a presença de (*loops*) ou laços de ordem três pode ser pelo coeficiente de clustering (*clustering coefficient*), ou cc .

Em redes que não tem pesos, o coeficiente de clustering cc_i (Eq. 2.3) de um nó i , é descrito como a relação entre a quantidade de ligações e_i entre os vértices da vizinhança e o total de ligações que poderia existir entre eles, ou seja, $(k_i(k_i - 1)/2)$. O termo final pode ser

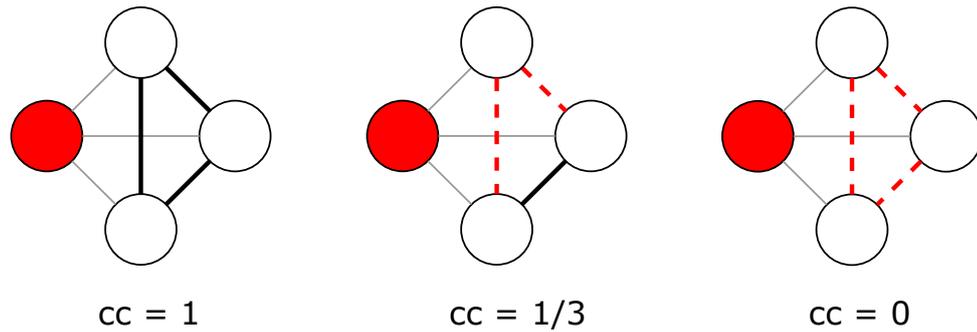


Figura 2 – Exemplos de coeficiente de clustering calculado para o nó em destaque. As linhas tracejadas representam a falta de conexão entre vizinhos.

Fonte: Elaborada pelo autor.

compreendido como sendo a multiplicação da número de vizinhos (k_i) pelo grau máximo ($k_i - 1$) para cada um deles. Como cada aresta é dividida entre pares de vizinhos, é importante dividir por dois.

$$cc_i = \frac{2e_i}{k_i(k_i - 1)} \quad (2.3)$$

A razão que identifica o coeficiente de clustering apresenta valores de 0, para o caso de não ter conexão entre os vizinhos, até o máximo de 1, caso todas as conexões entre os vizinhos estejam preenchidas, conforme a Figura 2.

Podemos também entender a medida cc como medida de centralidade, como se dois vértices fossem conectados unicamente pelo nó i , este vértice controlaria o fluxo de informação (NEWMAN, 2010). Note que cc_i assume os menores valores para os nós mais centrais, em oposição a outra medidas de centralidade.

Outra métrica geral usada para medir a ocorrência de triângulos em grafos é média do coeficiente de clustering de toda a rede, denominado de $\langle cc \rangle$, o qual calculamos conforme a equação,

$$\langle cc \rangle = \frac{1}{N} \sum_{i=1}^N cc_i. \quad (2.4)$$

2.1.4 Distância

Caminho em grafos é uma ordem sequencial de nós i_1, i_2, \dots, i_n em que cada um dos nós tenha uma aresta para o outro nó da sequência. O tamanho de uma distância ou caminho de i_1 a i_n é exatamente a quantidade de arestas pertencentes ao caminho, de valor $n - 1$ (em redes sem pesos). Em redes que apresentam pesos, a distância é a soma de cada peso das arestas.

Comunicação ou transporte em redes está relacionado com distanciamento entre vértices. Vejo o exemplo para redes geográficas: em redes de rodovias (Figura 3), a distância entre 2 pontos se relaciona com duração do transporte e também congestionamentos (MAY; MILNE, 2000). Os caminhos tem influência na transmissão de pacotes de dados pela Internet, e o atraso

(delay) ou perda de pacotes está relacionada aos caminhos, ou rotas selecionadas, e pode ocorrer congestionamentos e falhas de roteadores (TANENBAUM, 2003), que se assemelham aos encontros de rodovias. Desta forma, as distâncias e os caminhos são importantes para se analisar redes reais.

A distância contornada em um caminho é calculada pelo número de ligação entre dois nós (i, j) . Definimos o *comprimento do caminho mais curto*, como d_{ij} , escolhendo a menor das distâncias entre os possíveis caminhos que ligam i a j . As menores distâncias entre os pares de nós formam a matriz de distâncias D , que contém os elementos d_{ij} (menor caminho entre i e j). O valor máximo encontrado na matriz D é exatamente o *diâmetro* do grafo, denominamos de d_{max} , e a média de todos os elementos desta matriz é ℓ , o *menor caminho médio*, conforme equação

$$\ell = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}, \quad (2.5)$$

onde consideramos somente os componentes interligados, logo que, caso i e j não pertençam a um dado componente, teremos $d_{ij} = \infty$ e, desta forma, a soma divergiria.



Figura 3 – Rede de rodovias pelo Brasil. As rodovias representadas pelas arestas e as cidades representadas pelos pontos.

Fonte: Elaborada pelo autor.

2.1.5 Grau de proximidade closeness centrality

A medida de distância ℓ pode ser calculada para um determinado vértice, onde

$$\ell_i = \frac{1}{(N-1)} \sum_{j \neq i}^N d_{ij}, \quad (2.6)$$

assim, podemos definir a medida de proximidade (*closeness centrality*, em inglês), tomando o inverso de ℓ_i

$$C_i = \frac{1}{\ell_i} = \frac{N}{\sum_{j \neq i}^N d_{ij}}. \quad (2.7)$$

e d_{ij} é o percurso mais rápido entre os agentes i e j , e N é o quantidade de membros do grafo.

2.1.6 Betweenness centrality

A medida *betweenness centrality* serve para medir quanto um nó (aresta) u se localiza nos caminhos entre outros nós (FREEMAN, 1977). É uma medida que quantifica o nível de tráfego passando por u , levando em conta os menores caminhos. Podemos exemplificar por mensagens trocadas em uma rede social, supomos que os pares de nós (usuários) enviam mensagens com a mesma probabilidade no tempo, e as mensagens seguem os menores caminhos. Portanto, depois de muito tempo, a quantidade de mensagens que passam por cada nó é relacionada a quantidade de menores caminhos que se localizam sobre o nó. *Betweenness* é equacionado por

$$B_u = \sum_{ij} \frac{\sigma(i, u, j)}{\sigma(i, j)}, \quad (2.8)$$

$\sigma(i, u, j)$ fica sendo o somatório de caminhos curtos entre i e j que incluem o nó (aresta) u , onde $\sigma(i, j)$ é somatório de caminhos curtos entre i e j . A soma é feita sobre todos i, j . Já a medida global para uma rede é calculada pela média da medida *betweenness centrality* de todos os nós (arestas) u , ou seja,

$$B = \frac{1}{N} \sum_u^N B_u. \quad (2.9)$$

nós com valor alto de *betweenness centrality* podem influenciar bastante a rede, pelo fato de terem controle das informações que trafegam entre os outros nós, e removê-los pode prejudicar a comunicação entre os nós.

2.1.7 Assortatividade

Muitas redes reais apresentam correlação de grau pois a probabilidade de um nó com grau k ser ligado a outro nó de grau k' depende de k (BOCCALETTI *et al.*, 2006). Para isso, podemos usar a probabilidade $P(k'|k)$, que é definida pela probabilidade de uma aresta de um vértice de grau k se conectar com um vértice de grau k' . Desta forma, descrevemos esta probabilidade:

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle} \quad (2.10)$$

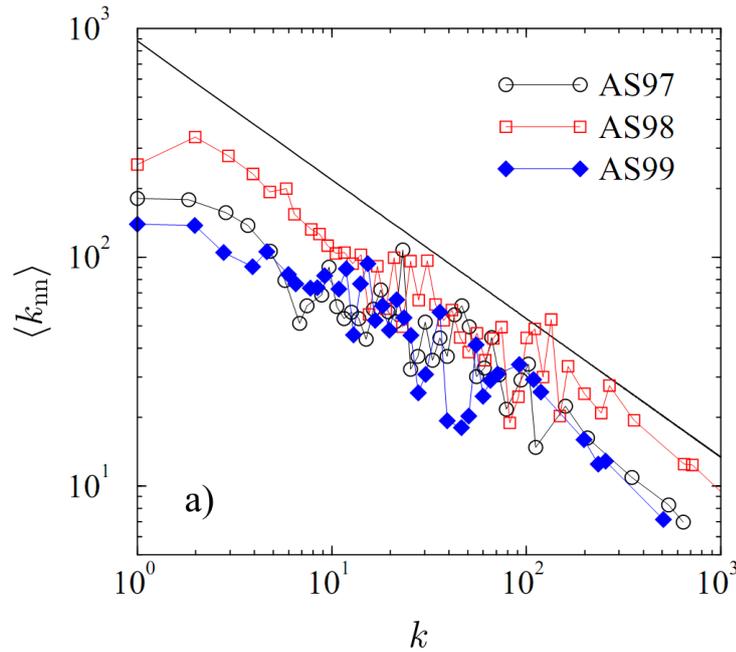


Figura 4 – $\langle k_{nn} \rangle$ da Internet entre 97 e 99.

Fonte: Adaptada de Vázquez, Pastor-Satorras e Vespignani (2002).

Tabela 1 – Exemplos de redes e suas Assortatividades.

Rede	N	r	Referências
Colaboração em física	52.909	0,363	(NEWMAN, 2001)
Colaboração em biologia	1.520.251	0,127	(NEWMAN, 2001)
Colaboração em matemática	253.339	0,120	(GROSSMAN; ION, 1995)
Colaboração de atores	449.913	0,208	(WATTS; STROGATZ, 1998)
Diretores de empresas	7.673	0,276	(DAVIS; YOO; BAKER, 2003)
Internet	10.697	-0,189	(CHEN <i>et al.</i> , 2002)
World Wide Web	269.504	-0,065	(BARABÁSI; ALBERT, 1999)
Interação de proteínas	2.115	-0,156	(JEONG <i>et al.</i> , 2001)
Rede neural	307	-0,163	(WATTS; STROGATZ, 1998)
Cadeias alimentares	92	-0,276	(MARTINEZ, 1991)

A média da conectividade dos vizinhos que estejam mais próximos de nós com grau k , é $\langle k_{nn} \rangle$, calculada assim:

$$\langle k_{nn} \rangle = \sum_{k'} k' P(k'|k). \quad (2.11)$$

Se $\langle k_{nn} \rangle$ é função que cresce com k , nós de alta conectividade se ligam entre si, e estes grafos são classificados como assortativos, e se $\langle k_{nn} \rangle$ decresce com k , nós de alta conectividade se ligam com nós que têm baixa conectividade, estes grafos são classificados como disassortativos. Se não existir correlação, $\langle k_{nn} \rangle$ não depende de k (NEWMAN, 2002).

Na Figura 4, vemos a relação da métrica $\langle k_{nn} \rangle$ com o grau k da Internet, que segue a relação $\langle k_{nn} \rangle \sim k^{-\nu}$, para $\nu = 0,5$. O valor obtido mostra que nós de alta conectividade provavelmente se ligam a nós de menor conectividade, ou seja, o grafo da Internet é disassortativo.

Outro método de localizar a correlação de k é considerar o coeficiente proposto por Pearson para correlação (NEWMAN, 2002):

$$x = \frac{\sum_{ij}(a_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij}(k_i \delta_{ij} - k_i k_j / 2m) k_i k_j} \quad (2.12)$$

onde a_{ij} é o membro da matriz A , k_i é o grau, m é quantidade de ligações e δ_{ij} é delta de Kronecker, ou seja, $\delta_{ij} = 1$ se $i = j$, e 0 caso diferente. Quando $x > 0$, a rede é assortativa, quando $x < 0$, a rede é dita disassortativa e quando $x = 0$ não existe correlação entre a conectividade dos nós.

No caso de redes assortativas os nós com conectividade similar tendem a se ligarem entre si, já nas disassortativas os nós de alto grau tendem a se ligarem a nós de grau baixo (NEWMAN, 2002). Na tabela 1, observamos que algumas redes são assortativas, como exemplo, a rede de colaboração de artigos de física, e outras são disassortativas, como exemplo, a rede de proteínas.

2.1.8 Grau médio de vizinhos

Calculamos o grau médio de vizinhos para um determinado vértice i , r_i (BOCCALETTI *et al.*, 2006), como a seguir,

$$r_i = \frac{1}{k_i} \sum_j^N A_{ij} k_j, \quad (2.13)$$

onde A_{ij} é o membro da matriz A , e k_j é o grau do nó do vizinho j . Definiremos uma medida geral, para caracterizar a estrutura de rede, por r , que é a média da conectividade média dos vizinhos de todos nós da rede, ou seja,

$$r = \frac{1}{N} \sum_i^N r_i. \quad (2.14)$$

Tem sido verificado que o grau médio de vizinhos é relacionado à propagação de epidemias em redes (BARRAT; BARTHLEMY; VESPIGNANI, 2008).

2.1.9 Page Rank

Um determinado vértice recebe um número de visitas quando um agente percorre a rede sem uma rota preferencial, e este número pode ser utilizado para quantificar a centralidade do vértice. Neste caso, uma medida possível é o PageRank (BRIN; PAGE, 1998) da Google, o qual é calculado da seguinte forma:

$$\pi^T = \pi^T G, \quad (2.15)$$

G é a matriz Google, descrita a seguir,

$$G = \kappa \left(P + \frac{ae^T}{N} \right) + \frac{(1 - \kappa)}{N} ee^T, \quad (2.16)$$

e a é o vetor binário chamado de vetor de nós pendurados ou nós sem conexões de saída (a_i é igual a 1, se i é um nó pendurado, e 0 em caso contrário), e é um vetor de uns com comprimento

N , e P é matriz que contém as probabilidades de transição da respectiva rede ($P(i, j) = \frac{1}{\sum_j A_{ij}}$, onde A é a adjacência). O índice π_i de uma dada página depende dos índices π_j das demais páginas que estão conectadas.

2.1.10 Centralidade por autovetor

Tal medida leva em consideração a centralidade de cada vértice como a soma dos valores de centralidade dos vértices que ele está ligado. Autovetor como medida de centralidade é definido pelo autovetor associado ao maior autovalor de A . Da seguinte forma,

$$x_i = \kappa^{-1} \sum_j A_{ij} x_j, \quad (2.17)$$

ou matricialmente, $A\mathbf{x} = \kappa\mathbf{x}$, onde \mathbf{x} é o autovetor principal (NEWMAN, 2010) e κ é o maior autovalor.

2.1.11 Coreness

A decomposição *k-core* ou *k-shell* sub-dividi uma rede em estruturas hierárquicas, e associa um índice para cada vértice i , o $k_c(i)$, de modo que $k_c(i) = k$ se i pertencer a (k)-core mas não pertencer a ($k + 1$)-core (SEIDMAN, 1983). Podemos dizer que os vértices com baixos valores de k_c estão localizados na periferia da rede, enquanto que os nós que apresentam altos valores dessa medida, são os mais centrais. Os hubs tendem a se ligarem concentrando o fluxo e formando um centro, muito parecidos com as ruas do centro da cidade, enquanto que a periferia recebe poucas "ruas"(arestas), por isso, os k_c baixos geralmente apontam grupos de vértices isolados na extremidade da rede. Recentemente adotou-se essa medida para detectar os principais propagadores em redes, conforme (KITSACK *et al.*, 2010).

2.2 Modelos de redes complexas

Após vermos as medidas para caracterizar estrutura, veremos que cientistas de teoria de redes desenvolveram modelos de geração de redes para prever a evolução da estrutura dos sistemas complexos. Estes modelos nos permite executar quatro tarefas: (i) prever: permite descrever como a organização do sistema está evoluindo no tempo; (ii) simular: auxilia a análise dos processos / dinâmicas, como exemplo, disseminação de epidemias, podendo usar topologias de redes complexas, como exemplo, a Internet; (iii) controlar: reproduz condições que corresponde a realidade mesmo nos ambientes virtuais e o monitora o comportamento do sistema sob uma dinâmica, como exemplo, modificar a largura de banda na Internet; e (iv) estudar falhas: verificar efeitos de remover arestas e nós da rede, como exemplo, a Internet, quando roteadores falham e geram tráfego distribuído e causa redução do desempenho. Desta forma, controlar, prever e o desenvolver mecanismos para deixar sistemas complexos com mais robustez pode ser realizada de uma forma eficiente e de baixo custo apenas através da modelagem.

Ínúmeros modelos de redes têm se desenvolvido para se estudar os incontáveis sistemas complexos (BOCCALETTI *et al.*, 2006). Entre os modelos, os mais descritos na literatura, e normalmente os mais usados, são explicados na sequência.

2.2.1 Redes aleatórias

No ano de 1959, matemáticos da Hungria, Alfred Rényi e Paul Erdős, utilizaram grafos como sendo elementos estocásticos, em vez de analisar de forma determinística como já era feito pela matemática discreta e pela sociologia. Neste modelo, chamado de *ER*, os grafos são construídos a partir de um conjunto de N nós não ligados entre si. Na sequência, o algoritmo seleciona 2 nós aleatoriamente e os conecta com probabilidade p , e cada par de nós é escolhido uma única vez. No final, todas as conexões terão a mesma probabilidade de ocorrer, isto é, o grafo gerado terá uma organização muito homogênea.

A distribuição de grau $P(k)$ possui valor característico determinado pelo grau médio $\langle k \rangle = p(N - 1)$, pois p é a probabilidade de definir a relação de pretendentes a formar conexão. Para altos valores de N e baixos de p , a distribuição aproxima da distribuição de Poisson (Figura 5(b)), por causa da lei de eventos raros do processo de Bernoulli (ROSS, 2007), ou seja,

$$p(k) = \lim_{N \rightarrow \infty, p \rightarrow 0} \frac{N!}{(N-k)!k!} p^k (1-p)^{N-k} \simeq \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}, \quad (2.18)$$

e $\langle k \rangle = p(N - 1)$ é o valor médio da conectividade.

Muito diferente da grade regular, o modelo *ER* mostra uma quebra de simetria deixando o caminho médio ℓ entre os nós (Eq. 2.5) pequeno se comparado ao tamanho da rede, sendo da

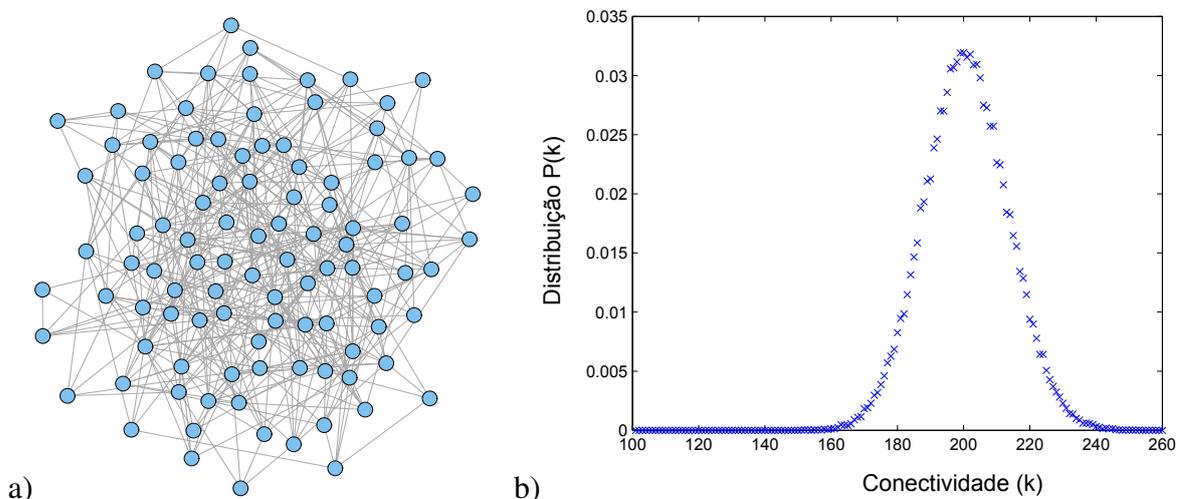


Figura 5 – (a) Rede aleatória *ER*. (b) Distribuição de grau de rede de mil nós, usando probabilidade $p = 0,2$. Média de várias simulações.

Fonte: Elaborada pelo autor.

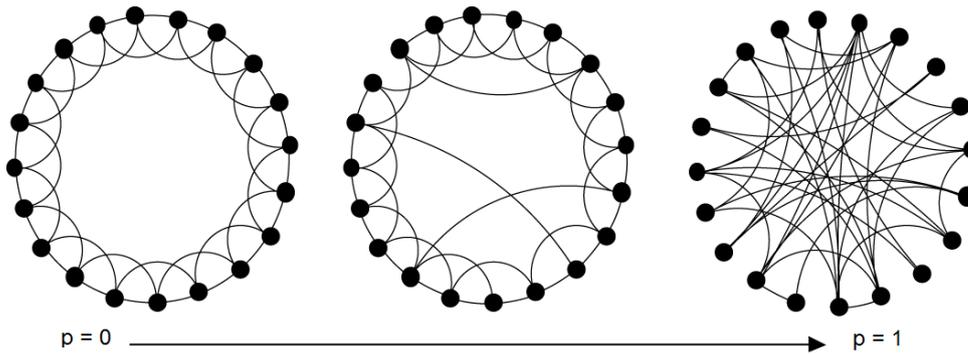


Figura 6 – Exemplos de redes variando p no modelo mundo pequeno com grau médio $\langle k \rangle = 4$.

Fonte: Adaptada de [Watts e Strogatz \(1998\)](#).

ordem de $\ln(N)/\ln(\langle k \rangle)$, e este caminho diminui quando a rede se torna esparsa para N muito maior que o $\langle k \rangle$.

2.2.2 Modelo Mundo Pequeno

No ano de 1998, nas Universidade de Cornell e Columbia, os cientistas observaram que algumas redes reais, como a de neurônios do *Caenorhabditis elegans* e a de distribuição de energia dos EUA a presença de caminhos fechados (*loops*) de ordem 3 muito maior do que nas redes aleatórias já estudadas e de mesmo número de nós e conexões ([WATTS; STROGATZ, 1998](#)). Assim acendeu o alerta que as redes reais não são tão aleatórias como se pensava, mas seguem leis que controlam a estrutura. Em cima desta descoberta, Watts e Strogatz sugerem um modelo alternativo a redes aleatórias, chamado de modelo mundo pequeno ou *small-world* (devido ao fenômeno que havia sido descoberto por Milgram S.). Neste modelo de geração de

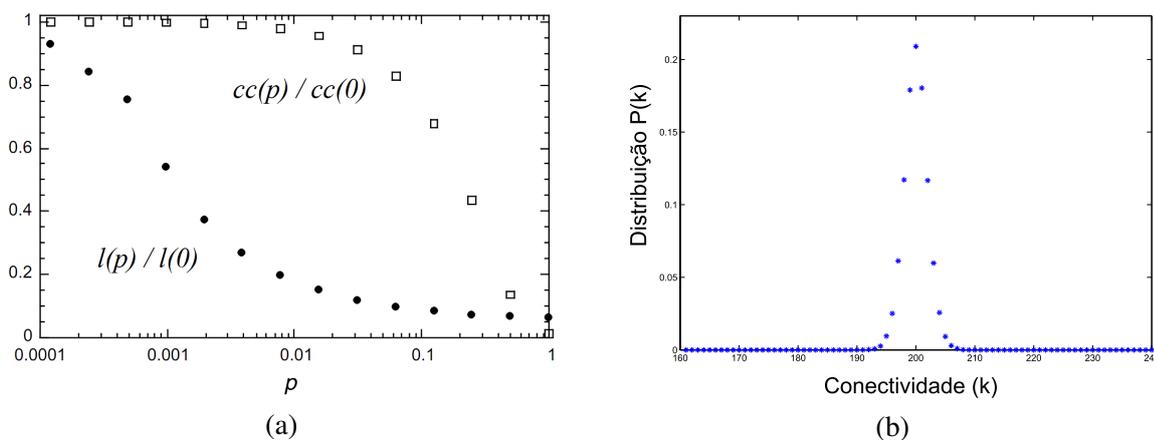


Figura 7 – (a) Gráfico da medida de coeficiente de clustering cc e do comprimento médio ℓ usando diferentes limites de probabilidade p aplicados ao modelo WS. (b) Distribuição de grau para um grafo com $N, 1000, \langle k \rangle, 200$, usando $p = 0,01$.

Fonte: Adaptada de [Watts e Strogatz \(1998\)](#).

rede, chamado de *WS*, inicia com grade regular constituída N nós interligados com k vizinhos nas duas direções, o que totaliza $2k$ conexões, onde $N \gg k \gg \log(N) \gg 1$. No passo seguinte, cada conexão é reconectada de forma aleatória com probabilidade p , que insere um caráter aleatório ao grafo. Para $p = 0$ o grafo é totalmente regular e para $p = 1$, o grafo é aleatório. Desta forma, este modelo fica entre a total regularidade e a completa aleatoriedade. Vemos o *small-world* aparecer em $p > 0,01$ (WATTS; STROGATZ, 1998). Isto pode ser visto na Figura 7(a), a variação do coeficiente de clustering cc e do média dos caminhos mais curtos ℓ , comparando com os valores iniciais de $p = 0$. Após valores de $p > 0,01$ é que se observa o efeito mundo pequeno, e abaixo deste valor o grafo se aproxima de uma grade regular, e acima temos uma queda brusca do valor de cc e também valores baixos de ℓ . Isto ocorre por causa do rearranjo das arestas, logo que os nós, que antes ficavam longe da vizinhança, podem agora dividir novas conexões, o que diminui a caminhada entre estes nós recém ligados. A distribuição de grau destes grafos é mostrada na Figura 7(b), onde os nós apresentam um número característico de ligações, como no modelo Erdős e Rényi.

2.2.3 Redes livre de escala

As redes sociais apresentarem propriedades grafos aleatórios como o efeito de mundo pequeno, como vimos anteriormente, mas estas não apresentam distribuição de grau de Poisson. Assim, os modelos *WS* e *ER* não são muito bons para representar a estrutura destas redes.

No ano de 1999, Réka Albert e Albert-László Barabási (BARABÁSI; ALBERT, 1999), começaram analisar a estrutura da *World Wide Web* através de um *web crawler*. Desta forma, eles descobriram a estrutura das ligações entre os sites da WWW e observaram que, além do fenômeno *small-world* ($\ell \approx 11$), a distribuição de grau não era aleatória, mas da forma *livre de escala* (*scale-free*), que segue $P(k) \sim k^{-\gamma}$ (ALBERT; JEONG; BARABÁSI, 1999). Este achado causou certo espanto, pois se esperava que a WWW fosse estruturada aleatoriamente, pois não tinha um comando central que ía definindo suas ligações. Porém, neste ano, esta distribuição tinha sido descoberta na rede da Internet pelos irmãos Faloutsos (FALOUTSOS; FALOUTSOS; FALOUTSOS, 1999), embora não tenha despertado grande impacto. Isto é, várias evidências mostravam que a estrutura da maioria dos sistemas complexos não era aleatória, mas surgem conforme conexão preferencial, o que foi observado também nas redes de colaboração de cientistas (BARABÁSI *et al.*, 2002), redes de interação de proteínas (JEONG *et al.*, 2001) e redes metabólicas de 43 organismos diferentes (JEONG *et al.*, 2000), sendo que nas redes metabólicas verifica-se que o número de ligações se relaciona com a letalidade das proteínas. Isto é, a distribuição de grau está relacionada às funções do sistema complexo.

Vários sistemas complexos apresentam distribuições da forma livre de escala, e o que era previsto por Rényi e Erdős não se observa na natureza. Desta forma, a pesquisa de Albert e Barabási derrubou a aleatoriedade e mostrou que existem leis que controlam a organização das redes do mundo real. Elas são idênticas, embora autores tenham causado confusão por

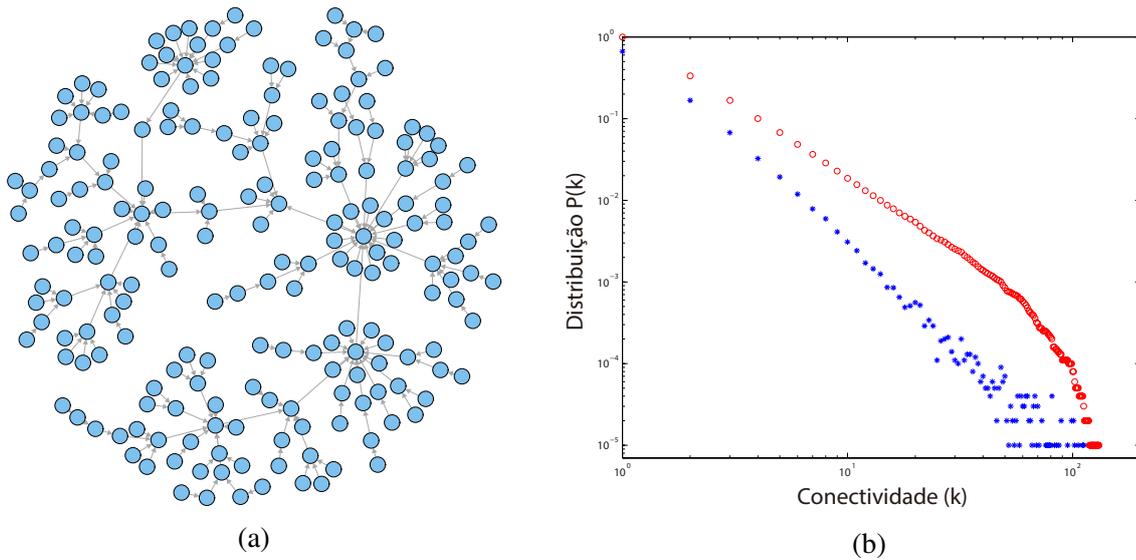


Figura 8 – (a) Rede do tipo livre de escala criada através do modelo *SF*. (b) Gráfico com escala logarítmica para distribuição de grau (em azul) e da distribuição cumulativa (em vermelho) para grafo de mil nós.

Fonte: Elaborada pelo autor.

diferenciá-las (NEWMAN, 2005). A distribuição de Pareto foi criada por V. Pareto (PARETO, 1935) no início do XX, e demonstrou que determinados fenômenos da economia, como em física, podem matematicamente ser modelados. Pareto observou que em muitos fenômenos econômicos e físicos 80% das consequências eram geradas pelos 20% das causas, e ficou conhecido como regra 80/20, uma curva decrescente e sem valor característico, descrita por um expoente negativo (livre de escala ou lei de potência). Desta forma, a economia não é gerida aleatoriamente, mas segue leis que controlam o comportamento. Deste modo, o tamanho das cidades, o diâmetro de crateras da Lua, distribuição de riqueza e a quantidade de citações de artigos, obedecem uma distribuição do tipo livre de escala, com ausência de média e desvio que caracterizam uma escala típica (como fenômenos aleatórios, escala característica, observe a Figura 5(b)).

Por causa desta falta de escala em redes reais, Albert e Barabási criaram um modelo de geração, que cria grafos livres de escala (BARABÁSI; ALBERT, 1999), que chamaremos de modelo *SF*, e que se baseia da seguinte forma:

1. *Aumento*: Começa com número de vértices N_0 , e a cada passo adiciona-se um nó de m ($m \leq N_0$) ligações que se conectam a nós já existentes.
2. *Conexão preferencial*: Este novo nó, que será adicionado ao grafo, tende a se ligar a nós mais conectados, isto é, a probabilidade de um nó j , ser selecionado é proporcional ao seu grau,

$$\mathcal{P}_{i \rightarrow j}(n+1) = \frac{k_j(n)}{\sum_{u=-N_0+1}^n k_u(n)}, \quad (2.19)$$

e n é tempo / número de nós adicionados ao grafo.

Interessante notar os dois pasos de geração dos grafos livres de escala não estão no modelo aleatório de Rényi e Erdős ou no modelo *small-world* de Strogatz e Watts, pois não há inserção de novos nós e as ligações são formadas homoganeamente, tendo um grau característico (grau médio) na rede. Em contra partida, nos grafos gerados pelo modelo livre de escala, os nós mais interconectados tendem a receber mais ligações. Isto é chamado por “*o rico fica mais rico*”.

Redes do tipo livre de escala são criadas por um número pequeno de nós de alta conexão, chamados de *hubs*, e por outra quantidade grande de nós de baixa conexão, onde vemos a falta de escala característica de distribuição. Na Figura 8 é apresenrado um grafo gerado pelo modelo de Albert e Barabási e a distribuição de grau para um grafo formado por mil nós, como também sua distribuição cumulativa $P_{cum}(k) = \int_k^\infty P(x)dx$.

2.2.4 Redes espaciais

O modelo proposto por Waxman (WAXMAN, 1988) considera que os nós estão uniformemente distribuídos em um quadrado de área unitária e cada par de nós está ligado de acordo com uma probabilidade que depende das distâncias, isto é,

$$p(i, j) = \eta \exp(-\eta d_{ij}), \quad (2.20)$$

onde η controla a conectividade e d_{ij} é a distância. O modelo gera grafos de distribuição exponencial, onde a probabilidade de um vértice ter grau diferente de $\langle k \rangle$ cai de forma exponencial.

O modelo proposto por Barthélemy (BARTHÉLEMY, 2003), cria grafos livre de escala e também espaciais. Considera-se uma estrutura de dimensão d , regular e de tamanho L , e tem 3 etapas. Primeiro, n_0 vértices iniciais são aleatoriamente escolhidos. Em seguida, um nó i inativo é selecionado aleatoriamente, e ligado a um vértice ativo j e de probabilidade

$$p(i, j) \propto \frac{k_j + 1}{\exp(d_{ij}/r_c)}, \quad (2.21)$$

onde k_j é o número das ligações de nó j , r_c é parâmetro da escala finito e d_{ij} é a distância Euclidiana entre os vértices i e j . Por fim, i fica ativo e os demais passos são refeitos até que todos os outros estejam ativos. O parâmetro r_c serve para controlar o coeficiente de clustering (WATTS, 1999) e o de assortatividade (NEWMAN, 2002).

PROPAGAÇÃO DE EPIDEMIAS E RUMORES

Podemos dizer que o primeiro modelo epidemiológico foi proposto por Bernoulli em 1760 (DIETZ; HEESTERBEEK, 2002). No entanto, somente no século XX, os modelos matemáticos foram sendo aprimorados para serem usados em situações reais. Os modelos simplificados levam em conta infecções que são imunizantes mas também as que não permitem imunidade, assim o indivíduo pode se reinfectar. Em resumo os indivíduos são suscetíveis à infecção, infectados ou recuperados (eram infectados e depois ficaram imunizados). A seguir, descrevemos os modelos epidêmicos considerados nesse trabalho.

3.1 Modelo *SIR*

No modelo *SIR* (suscetível-infectado-recuperado) temos três classes, ou compartimentos, de indivíduos, que definiremos como **S**uscetível, **I**nfectado e **R**ecuperado. A questão que estamos interessados é como os indivíduos se deslocam de um compartimento para o outro. Para simplificar (ignora-se o aumento e diminuição da população: nascimentos e mortes), ficamos com as transições $I \rightarrow R$ e $S \rightarrow I$ (Figura 9). Modelando, o indivíduo que se desloca de *I* a *R* depende do tempo ele ficou na classe infectados. Cientistas muitas vezes fazem a hipótese simplificada de que a taxa de recuperação μ (inverso do período de infecção) seja constante, então temos equações simples.

Agora para ir de *S* para *I* depende da propagação da doença, e é determinada por três fatores: a permanência de infectados, a organização de contato da população e da probabilidade de se transmitir por contato. Para doença ser transmitida tem que haver contato entre os suscetíveis e os infectados. A probabilidade de isso acontecer é determinada pelas frações indivíduos nos compartimentos *S* e no *I*, bem como pela estrutura de contato inerente da população hospedeira. O diagrama de fluxo está na Figura 9. O código fonte, para o modelo de propagação *SIR*, pode ser observado na Subrotina 1, tendo como saída um vetor (curva gráfica) que descreve o comportamento do epidemia.

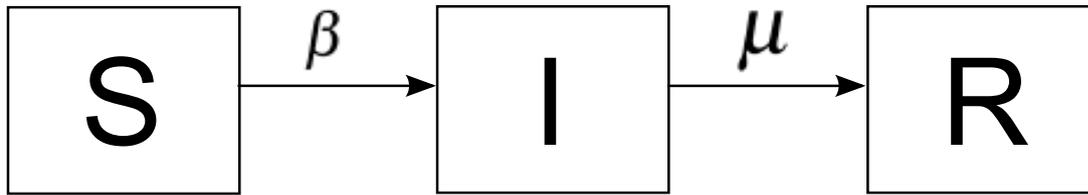


Figura 9 – Diagrama de fluxo do modelo de epidemia SIR.

Fonte: Elaborada pelo autor.

Código-fonte 1 – Subrotina para obter curva do modelo de epidemia SIR

```

1: function saida = SIR_i(i,A,N,R,c)
2:     beta = 0.3; mu = 1; steps = 60; saida = 0; curva = zeros(
3:         steps,1);
4:     for Nr=1:R
5:         % todos suscetiveis
6:         state = zeros(size(A,1),1,'single');
7:         % 0 suscetivel / 1 infectado / -1 recuperado
8:         lista = i; % vértice selecionado
9:         state(lista) = 1; % infectou o primeiro
10:        for f=1:steps
11:            lista = find(state == 1); % infectados
12:            for q=1:length(lista)
13:                atual = lista(q); % Propagador atual
14:                neig = find(A(atual,:) == 1);
15:                for s=1:length(neig) % vizinho
16:                    if(state(neig(s)) == 0) % suscetivel
17:                        if(rand() < beta)
18:                            state(neig(s)) = 1; % propagando
19:                        end
20:                    end
21:                end
22:            % Etapa de recuperacao
23:            for r=1:length(lista)
24:                if(rand() < mu)
25:                    state(lista(r)) = -1; % se recuperou
26:                end
27:            end
28:        % obtém fração da classe
  
```

```

29:             curva(f) = length(find(state == c))/N; % c == S/I/R
30:         end
31:         % Média dos resultados
32:         saida = saida + curva;
33:     end
34:     saida = max(saida/R);
35: end

```

3.1.1 Equações do modelo SIR

Para introduzir as equações do modelo, é mais fácil considerar uma "população fechada", sem dados demográficos (sem nascimentos, óbitos e migrações). Assumimos um padrão de conexões homogêneo, em que as complexidades que afetam o padrão de contatos são descartadas, resultando em βSI como termo de transmissão. Dada a premissa de que as probabilidades epidemiológicas subjacentes são constantes ao longo do tempo, obtemos as seguintes equações para o modelo *SIR*:

$$\frac{dS}{dt} = -\beta SI \quad (3.1)$$

$$\frac{dI}{dt} = \beta SI - \mu I \quad (3.2)$$

$$\frac{dR}{dt} = \mu I \quad (3.3)$$

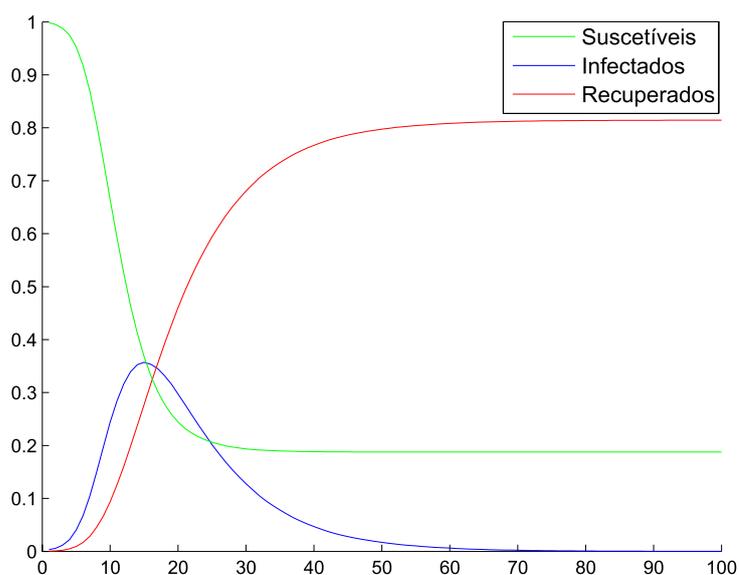


Figura 10 – Evolução do modelo *SIR*.

Fonte: Elaborada pelo autor.

O parâmetro μ é chamado de taxa de remoção ou recuperação, embora muitas vezes estamos mais interessados em seu recíproco ($1/\mu$), que determina o período médio de infecção. Note que epidemiologistas normalmente não descrevem a equação para a classe R , pois sabemos que $S + I + R = 1$, então tendo as classes S e I nos permitirá calcular as equações em função de R . Essas equações têm as condições iniciais $S(0) > 0, I(0) > 0$ e $R(0) = 0$.

Um exemplo da evolução de uma epidemia gerada a partir das equações do modelo SIR é apresentado na Figura 10. A fração de infectados cresce exponencialmente até um pico máximo, sendo que a partir daí, a fração de infectados decresce dada a alta taxa de recuperados.

3.2 Modelo S / S

Enquanto o modelo anterior permite imunidade, muitas outras doenças infecciosas não permitem imunidade, como a doença causada pelo rotavírus, infecções sexualmente transmissíveis, e também infecções bacterianas. De maneira simples, o modelo SIS é descrito por um par de equações diferenciais ordinárias acopladas:

$$\frac{dS}{dt} = \mu I - \beta SI \quad (3.4)$$

$$\frac{dI}{dt} = \beta SI - \mu I \quad (3.5)$$

Os parâmetros são os definidos como na seção anterior, mas com $S + I = 1$. Podemos, portanto, substituir $S = 1 - I$ na equação (3.5) e simplificá-la para obter

$$\frac{dI}{dt} = (\beta - \beta I - \mu)I = \beta I((1 - 1/R_0) - I) \quad (3.6)$$

onde, geralmente, $R_0 = \beta/\mu$ (BARRAT; BARTHLEMY; VESPIGNANI, 2008). O número de equilíbrio de infectados nesta população é obtido igualando a derivada a zero e resolvendo para I . Isto nos permite obter $I = (1 - 1/R_0)$, portanto, mais uma vez $S = 1/R_0$, e o equilíbrio será estável desde que $R_0 > 1$.

3.3 Modelo de rumor

Dinâmicas de propagação de rumor são, em alguns aspectos semelhantes à de epidemias (DALEY; GANI; GANI, 2001; CASTELLANO; FORTUNATO; LORETO, 2009). A propagação de rumor é modelada considerando que vértices são propagadores, *stiflers* ou ignorantes. Propagadores são aqueles indivíduos que conhecem o rumor e querem espalhá-lo para os ignorantes, enquanto que *stiflers* são aqueles que conhecem o rumor, mas não estão mais interessados nas informações. A principal diferença entre a propagação de rumores e epidemias é que um vértice propagador transforma-se em um *stifler* por um processo que envolve contatos,

enquanto um nó infectado torna-se recuperado por um processo espontâneo. A classe de ignorantes (I), propagadores (S), e *Stiflers* (R) são definidos de tal modo que $I + S + R = 1$, sendo as equações que regem a dinâmica:

$$\frac{dS}{dt} = -\beta SI \quad (3.7)$$

$$\frac{dI}{dt} = \beta SI - \mu I \quad (3.8)$$

$$\frac{dR}{dt} = \mu I. \quad (3.9)$$

O processo de difusão do rumor inicia com um propagador e $N - 1$ ignorantes, onde N é o número de nós na rede. Com taxa λ , propagadores difundem o rumor para seus vizinhos que sejam ignorantes. Com taxa δ , caso o propagador entre em contato com outro da mesma classe, o propagador se torna um *stifler*. Modelo proposto por Thompson e Maki e que chamaremos de **MT** (CASTELLANO; FORTUNATO; LORETO, 2009). Já no modelo proposto por Kendall Daley e que chamaremos de modelo **DK**, os dois propagadores que se encontram viram *stiflers* com taxa de λ (CASTELLANO; FORTUNATO; LORETO, 2009). Simulações da dinâmica de Monte Carlo para difusão de rumor podem ser de 2 formas. **CP**, processo de contato onde somente 1 vizinho do propagador é contatado, de forma aleatória, a cada etapa. No processo truncado (**TP**) os vizinhos de um propagador são contatados de forma aleatória até que todos eles sejam contatados ou o propagador se transforma em um *stifler*.

Os modelos de epidemias e rumores serão considerados nos próximos capítulos, quando estudaremos o problema da identificação dos propagadores mais influentes.

IDENTIFICAÇÃO DE PROPAGADORES INFLUENTES ATRAVÉS DE MINERAÇÃO DE DADOS

A propagação de epidemias depende da distância entre os vértices na rede. Assim sendo, como os vértices mais centrais são aqueles que mais facilmente acessam os outros vértices, através de troca de informações, temos que os nós mais centrais deve ser os propagadores mais influentes. Porém, quantificar esta centralidade é uma questão ainda em aberto na teoria das redes complexas. Diferentes medidas tem sido desenvolvidas, cada uma baseada em um conceito específico, tal como caminhadas aleatórias e caminhos mais curtos. Neste capítulo, propomos uma abordagem multivariada que leva em conta conceitos de agrupamento de dados para identificar os propagadores mais influentes. Simulações numéricas mostram que esta estratégia permite a identificação dos propagadores mais importantes com acurácia maior do que métodos tradicionais baseados em uma única medida de centralidade.

Uma das métricas de centralidade mais usadas foi o grau. No caso desta medida, um vértice altamente conectado pode propagar sua informação através de um grande número de nós vizinhos. Contudo, análises recentes tem sugerido que o número de conexões não tem um impacto fundamental na propagação de doenças (KITSAK *et al.*, 2010). De fato, se um nó fortemente conectado está na periferia de uma rede (identificado de forma visual), sua capacidade de propagação através do centro da rede será mínima (KITSAK *et al.*, 2010). Por outro lado, tem sido observado que vértices no núcleo da rede, identificados pela decomposição *k-shell*, desempenham um papel fundamental na propagação de doenças (KITSAK *et al.*, 2010). Apesar disso, tais nós não são os mais influentes propagadores de rumores (BORGE-HOLTHOEFER; MORENO, 2012). Também tem sido verificado que a centralidade *betweenness* é outra propriedade central mas que contribui pouco para a propagação de doenças, conforme observado por Kitsak *et al.* (2010). Ou seja, não há um consenso sobre qual a medida central mais adequada para identificar

os principais propagadores de doenças infecciosas e rumores.

Uma solução natural para superar esta limitação é considerar não uma única métrica para caracterizar a centralidade de um nó, mas sim utilizar um conjunto de medidas. Por exemplo, os pesquisadores [Erkol, Castellano e Radicchi \(2019\)](#) utilizaram 16 métodos baseados na estrutura da rede para apontar os nós mais influentes. Quanto mais centrais forem os descritores considerados, mais precisa será a caracterização de um nó central. Consideramos utilizar um conjunto de seis medidas mais comuns para quantificar o nível de centralidade de um nó, sendo elas o grau, *betweenness*, *closeness*, grau médio dos vizinhos, PageRank e *k-core*. Estas medidas estão relacionadas a diferentes conceitos, isto é, número de conexões, caminhos mais curtos, caminhadas aleatórias e decomposição *k-shell* ([NEWMAN, 2010](#)). Verificamos que o conjunto de nós com os mais altos valores dessas medidas são as mais influentes propagadores de doenças. Tais vértices foram identificados automaticamente considerando-se o algoritmo de maximização da esperança, que permite o agrupamento de vértices com característica semelhantes. Simulações numéricas mostraram que essa estratégia identifica os propagadores mais importantes com maior precisão do que os métodos baseados em grau e *k-core*.

As próximas seções fornecem as definições de como quantificar a propagação de epidemia e do conceito do algoritmo de maximização da esperança. Resultados e discussões são apresentados posteriormente. As definições das medidas de centralidade, utilizadas nesta análise, podem ser encontradas na Seção 2.1.

4.1 Propagação da epidemia

A fim de quantificar a capacidade de difusão de um determinado nó j , determinou-se o valor máximo da porcentagem de nós infectados, quando o processo de propagação iniciou-se em j . Quantificamos a capacidade de propagação de um nó j por

$$s(j) = \max\{i_j(t), t > 0\}, \quad j = 1, \dots, N, \quad (4.1)$$

onde $i_j(t)$ é a porcentagem de nós infectados em um tempo t após o processo de propagação ter iniciado no nó j (queremos aqui analisar o propagador inicial como o responsável pelo surto). Note-se que um tal pico da infecção, $s(j)$, está associada com a velocidade de propagação da doença, uma vez que permite a determinação da porcentagem de nós infectados antes do processo de recuperação dominar o processo de propagação.

4.2 Maximização da esperança

Métodos de agrupamento visam a identificar grupos de objetos que são semelhantes em termos de um conjunto de atributos. Métodos de agrupamento baseados em probabilidades têm sido considerados para encontrar a probabilidade de cada objeto pertencer a cada grupo ([THEODORIDIS; KOUTROUMBAS, 2003](#)). Estes métodos baseiam-se nos chamados modelos de

mistura finitos. Mais especificamente, uma função de densidade $p(x)$ pode ser modelada como combinação linear de outras função de densidade como $p(x) = \sum_{j=1}^k p(x|j)P_j$, onde $\sum_{j=1}^J P_j = 1$ e $\int_x p(x|j)dx = 1$. A mistura é este conjunto de k distribuições de probabilidades.

A Maximização da Esperança (ou EM, sigla em inglês para *Expectation Maximization*) é um algoritmo que calcula parâmetros de uma função de probabilidade multivariada com uma mistura de distribuições gaussianas (DEMPSTER; LAIRD; RUBIN, 1977). Esta abordagem baseia-se em duas etapas. A primeira, chamada passo E, consiste de uma estimativa da probabilidade $p(i, k)$ da amostra i que pertence à mistura k . O segundo passo, chamado passo M, estima os parâmetros da distribuição de probabilidades de cada classe pela maximização da probabilidade de as distribuições dadas os dados disponíveis. A grande vantagem desta abordagem é que determina o número de agrupamentos e elementos em cada grupo automaticamente por um procedimento de validação cruzada (WITTEN; FRANK, 2005). Informações adicionais sobre o método de EM pode ser encontradas em (WITTEN; FRANK, 2005; THEODORIDIS; KOUTROUMBAS, 2003; BISHOP *et al.*, 2006).

4.3 Conjunto de dados

A fim de analisar a influência da centralidade da rede sobre os processos de espalhamento, foram considerados as seguintes redes complexas: (a) rede de e-mails obtidas a partir de mensagens trocadas entre os usuários na *Universidade Rovira e Virgili*, Espanha (GUIMERA *et al.*, 2003); (b) redes do Facebook de faculdades e universidades obtidas em um único dia em setembro de 2005 (TRAUD; MUCHA; PORTER, 2011); e (c) rede de blogs de políticos (chamada *Political blogs*), de versão não dirigida, composto por hyperlinks entre weblogs obtidos durante o período de dois meses que antecederam as eleições presidenciais de 2004 nos Estados Unidos (ADAMIC; GLANCE, 2005). Para as redes do Facebook, mostramos apenas os resultados relativos às redes denominadas Smith60 e Santa74, como indicado no artigo original (TRAUD; MUCHA; PORTER, 2011), uma vez que os resultados para outras redes são semelhantes. Também consideramos um grafo em árvore, que é construído por meio da atribuição do grau de um nó como o número do nível mais um. Mais especificamente, nesta árvore, o nó central tem duas ligações, e seus vizinhos, no segundo nível, têm três ligações, e assim por diante, de modo que os hubs são colocados na fronteira da rede. Nós usamos somente o maior componente de cada rede considerada.

A Tabela 2 mostra as propriedades topológicas dessas redes, em que N é o número de nós, $\langle k \rangle$ é o grau médio, $\langle cc \rangle$ é o coeficiente de aglomeração, x é o coeficiente de assortatividade, e $\langle l \rangle$ é o menor caminho médio. Todas estas medidas estão descritas na Seção 2.1.

Tabela 2 – Propriedades de rede sem escala: número de nós (N), grau médio ($\langle k \rangle$), média do coeficiente de clustering ($\langle cc \rangle$), coeficiente de assortatividade (x) e médias dos caminhos mais curtos ($\langle l \rangle$).

Rede	N	$\langle k \rangle$	$\langle cc \rangle$	x	$\langle l \rangle$
Santa74	3578	84.82	0.263	0.070	2.486
Smith60	2970	65.41	0.283	0.044	2.498
Email	1133	9.61	0.220	0.078	3.606
Political blogs	1222	31.24	0.320	-0.221	2.738
Artificial tree	1000	1.998	0.0	-0.897	7.238

4.4 Resultados e discussão

A fim de caracterizar a centralidade de cada nó, consideramos o seguinte conjunto de medidas de centralidade: (i) grau do nó (k_i), (ii) *betweenness centrality* (b_i), (iii) *closeness centrality* (c_i), (iv) conectividade média dos vizinhos (r_i), (v) PageRank (pr_i), e (vi) k -core (kc_i). Estas medidas foram calculadas para cada nó. Notem que, enquanto que o grau de nó é um valor inteiro maior do que um, o PageRank é um número real na intervalo $[0 - 1]$. Para evitar esses efeitos de escala durante o agrupamento de dados, adotamos um procedimento de normalização. Levamos em conta o chamado escalonamento *softmax*, que tem sido amplamente utilizada na mineração de dados (THEODORIDIS; KOUTROUMBAS, 2003). Softmax é um tipo de função que limita a medida no intervalo $[0, 1]$. Portanto, cada medida μ é normalizada como:

$$z_i = \frac{1}{1 + e^{w_i}}, \quad (4.2)$$

onde

$$w_i = \frac{\mu_i - \langle \mu \rangle}{\sigma_\mu}, \quad (4.3)$$

e $\langle \mu \rangle$ and σ_μ são a média e desvio padrão da medida μ calculada sobre todos os nós, respectivamente. Após este procedimento de padronização, foi realizado um agrupamento de dados, considerando-se o algoritmo EM. Cada nó foi representada por um vector aleatório composto por seis medidas, isto é, $v_i = \{k_i, b_i, c_i, r_i, pr_i, kc_i\}$. O agrupamento foi baseado na semelhança de tais vectores. Note-se que este passo foi realizado sem conhecimento prévio do número de agrupamentos e elementos em cada classe.

A seguir, simulamos uma epidemia. A capacidade de propagação de cada vértice foi quantificada pela porcentagem máxima de nós infectados (Eq. 4.1). Observa-se que a propagação é iniciada num único vértice, que foi a semente do processo de difusão. De modo a obter resultados estatisticamente significativos, foram realizadas $S = 100$ simulações para cada nó, sendo que usamos a média dos resultados (fração máxima de infectados). Para cada grupo, identificado previamente em termos de medidas de centralidade, pelo EM, determinou-se a distribuição de probabilidade do percentual máximo de nós infectados. Figura 11(a) mostra tais distribuições para cada cluster na rede Santa74 do Facebook. Para uma melhor visualização, as distribuições

foram estimadas considerando-se janelas de Parzen com um kernel gaussiano (PARZEN, 1962). Em todos os casos, as distribuições apresentam médias diferentes, o que sugere que a capacidade propagação depende das propriedades centrais. Note-se que os grupos foram obtidos a partir de um conjunto de medida central, sem ter em conta as taxas de propagação. Portanto, cada distribuição da Figura 11(a) corresponde aos nós com medidas de centralidade semelhantes.

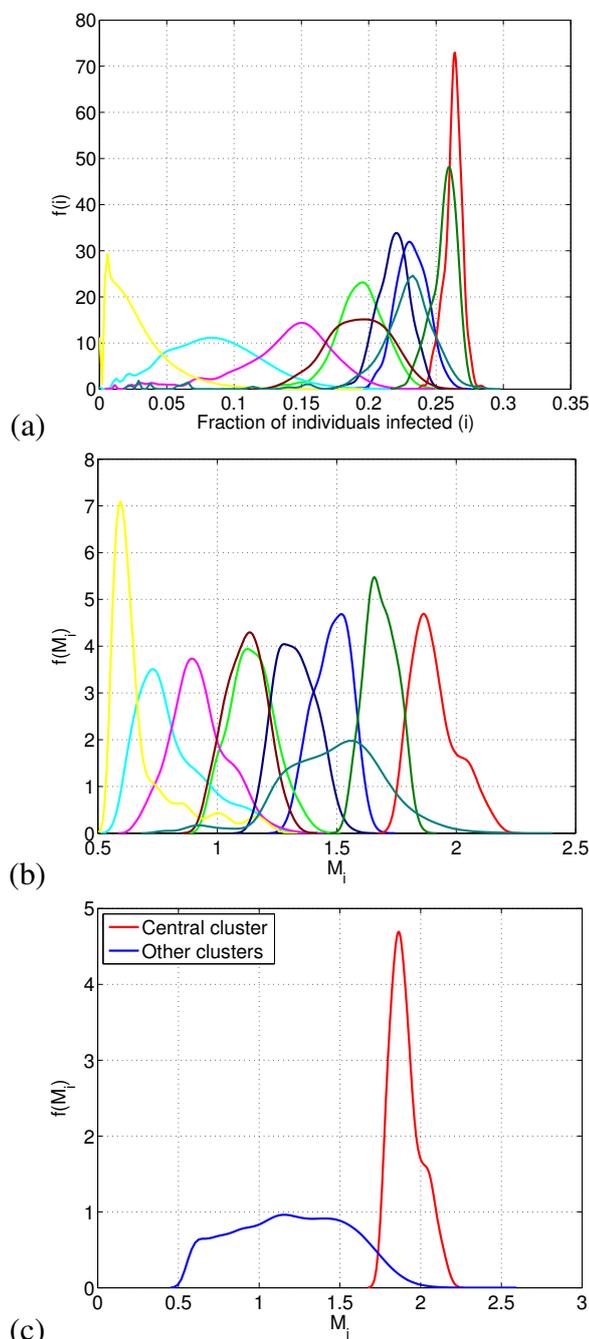


Figura 11 – Rede Santa74 do Facebook: (a) distribuições de probabilidades do percentual máximo de nós infectados no modelo SIR, (b) distribuições de índice central M_i e (c) distribuição de índice central M_i considerando-se apenas o cluster que contém os propagadores mais influentes e do cluster com o restante vértices.

Fonte: Elaborada pelo autor.

A Figura. 11(a) mostra que os propagadores mais influentes são os nós do cluster, cuja probabilidade de distribuição está em vermelho (no lado direito). A fim de verificar se os nós são os mais centrais, sugerimos uma definição multivariada da centralidade nó. Esta medida é definida como a norma do vector contendo as medidas de centralidade, isto é,

$$M_i = \|v_i\| = \sqrt{k_i^2 + b_i^2 + c_i^2 + n_i^2 + pr_i^2 + kc_i^2} \quad (4.4)$$

onde cada medida central foi padronizado pela Equação. 4.2. Note que $0 \leq M_i \leq 6$, uma vez que levamos em conta seis medidas. Portanto, os nós mais centrais pertencem ao cluster com maiores valores de M_i . Tais nós podem ser entendidos como valores atípicos, em termos de centralidade (COSTA *et al.*, 2009; COSTA; RODRIGUES, 2009). Fig. 11(b) mostra a distribuição de probabilidade de M_i para cada grupo identificado na rede Santa74 do Facebook. Em comparação com a distribuição da fracção dos indivíduos infectados mostrados na Figura. 11(a), podemos ver uma forte correspondência entre US M_i e a capacidade de propagação, portanto, o cluster que contém os nós com os mais altos valores de M_i têm os propagadores mais influentes na rede. A Figura 12 mostra a rede Santa74 e sua atribuição de cluster, representado por suas cores. Podemos ver que os nós do cluster com os maiores valores de M_i estão no centro da rede, sendo os principais propagadores. Esta correspondência entre uma posição central e a capacidade de propagação foi observada para as outras redes, por conseguinte, é possível identificar os propagadores mais influentes por propriedades centrais.

Para realizar a análise das redes restante, agrupamos os vértices em apenas dois clusters. Mais especificamente, para cada rede, o cluster que contém os nós mais centrais foi definido como aquele com o maior valor médio de M_i , enquanto que os clusters restantes foram agrupadas em um único cluster. A Figura 11 (c) mostra a distribuição de probabilidade do índice de centralidade M_i para rede Santa74 do Facebook considerando-se apenas dois clusters. Através da realização de um teste t de Student para comparação das duas distribuições, verificou-se que os valores médios da porcentagem máxima de nós infectados diferem significativamente no nível de 0.01 para todas as redes avaliadas. Assim, os dois grupos exibem propriedades centrais distintas em todas as redes consideradas.

Comparamos a capacidade de propagação dos vértices no cluster mais central com abordagens que levem em conta o grau de nó e o k -core. Nós consideramos um procedimento de imunização, que é realizado por vacinar todos os vértices no cluster com os maiores valores de M_i . A partir de agora, o processo propagação inicia-se em todos os outros vértices de uma vez. Para fins de comparação, os vértices com os maiores valores de grau ou k -core foram imunizados e as simulações foram realizadas. O número de vértices vacinados foi igual ao número de vértices no cluster mais central, a fim de termos uma comparação adequada. Uma imunização aleatória também foi considerada para fins de comparação. A Figura 13 mostra o valor máximo de vértices infectados para a rede de comunicação de emails, blogs políticos e as redes Santa74 e Smith60 do Facebook. A taxa de transmissão β foi variada de zero a um, em passos de $\Delta\beta = 0.05$. Podemos ver que, para todos os casos, a imunização dos nós no cluster mais central revelou ser a melhor

estratégia para diminuir o impacto do processo de propagação. Portanto, os nós nos principais aglomerados podem ser considerados os mais influentes propagadores da doença em todas as redes analisadas.

Ao contrário do observado antes (KITSACK *et al.*, 2010), a medida *k-core* não está fortemente associada com a capacidade de propagação, uma vez que a imunização de vértices com os mais altos valores de *k-core* tem resultados comparáveis com os que usam imunização aleatória. Além disso, podemos ver que para as redes Facebook, a imunização dos vértices no cluster mais central coincidiu com a imunização de hubs. Esta é uma indicação de que nestas redes, os principais hubs tendem a estar no núcleo principal da rede. Verificamos esta propriedade na rede de Barabási-Albert, no qual os hubs são os nós mais antigos na parte central da rede (por exemplo, (BOCCALETTI *et al.*, 2006)).

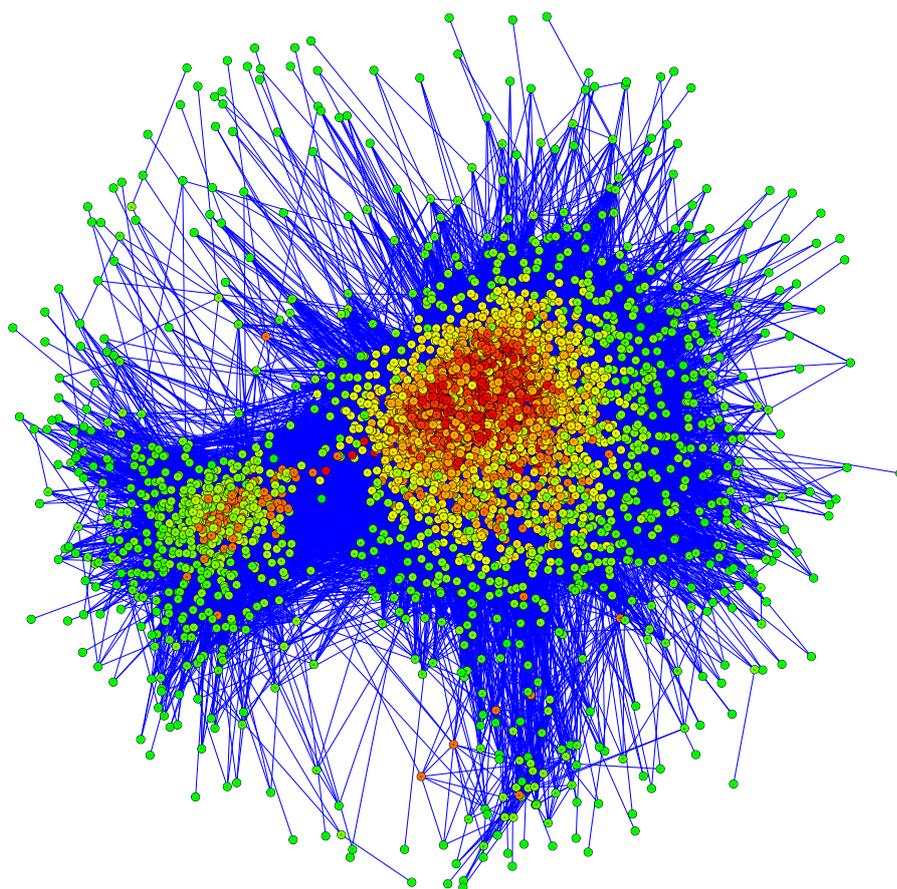


Figura 12 – Visualização da rede Santa74 do Facebook. Os nós na curva em vermelho pertencem ao cluster com os maiores valores de M_i . Esta visualização baseia-se num paradigma “força-dirigida” por Kawai e Kamada (KAMADA; KAWAI, 1989), disponível para Cytoscape (SHANNON *et al.*, 2003).

Fonte: Elaborada pelo autor.

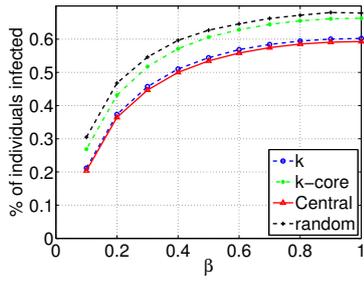
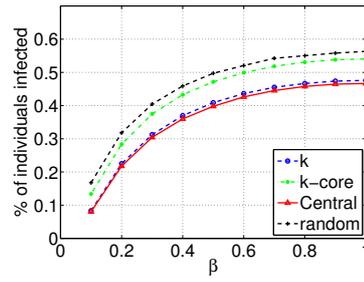
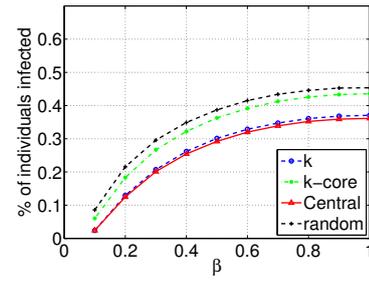
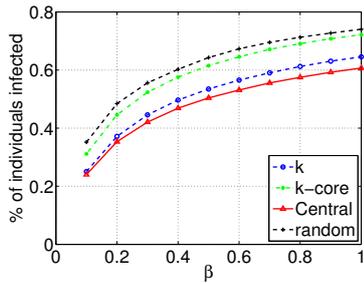
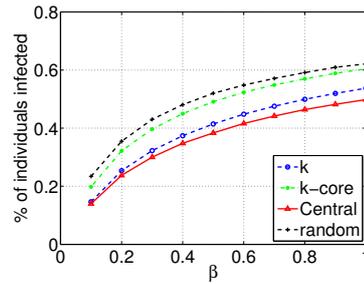
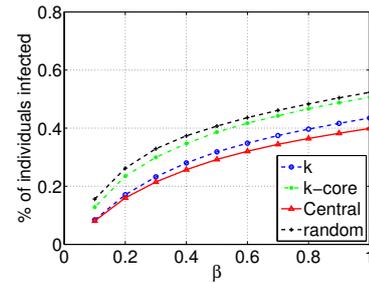
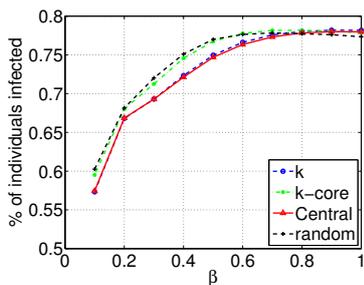
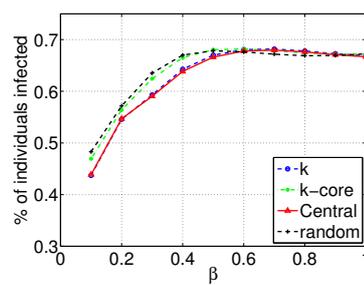
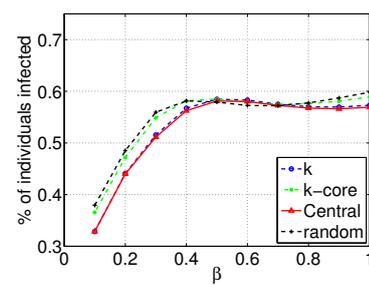
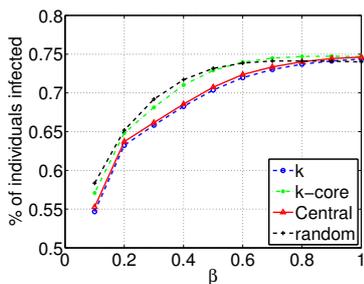
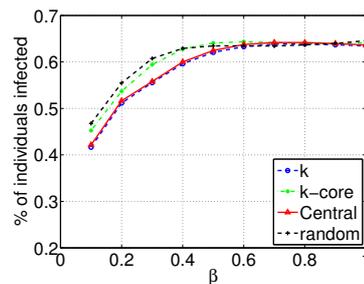
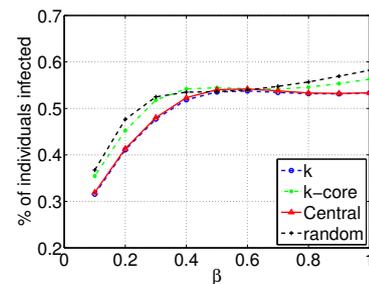
(a) Email network, $\mu = 0.3$.(b) Email network, $\mu = 0.6$.(c) Email network, $\mu = 1.0$.(d) Political blogs network, $\mu = 0.3$.(e) Political blogs network, $\mu = 0.6$.(f) Political blogs network, $\mu = 1.0$.(g) Facebook Smith60 network, $\mu = 0.3$.(h) Facebook Smith60 network, $\mu = 0.6$.(i) Facebook Smith60 network, $\mu = 1.0$.(j) Facebook Santa74 network, $\mu = 0.3$.(k) Facebook Santa74 network, $\mu = 0.6$.(l) Facebook Santa74 network, $\mu = 1.0$.

Figura 13 – Porcentagem máxima de indivíduos contaminados após procedimentos de imunização com base no grau, k -core, cluster mais central e aleatório para a rede de email (a) - (c), rede Political blogs (d) - (f), rede Smith60 do Facebook (g) - (i) e rede Santa74 do Facebook (j) - (l). Diferentes valores de taxa de difusão (β) e recuperação (μ) foram considerados.

Fonte: Elaborada pelo autor.

No que diz respeito as outras redes, embora a imunização dos nós dos aglomerados mais centrais produziram os melhores resultados, a sua diferença em relação ao grau de vacinação

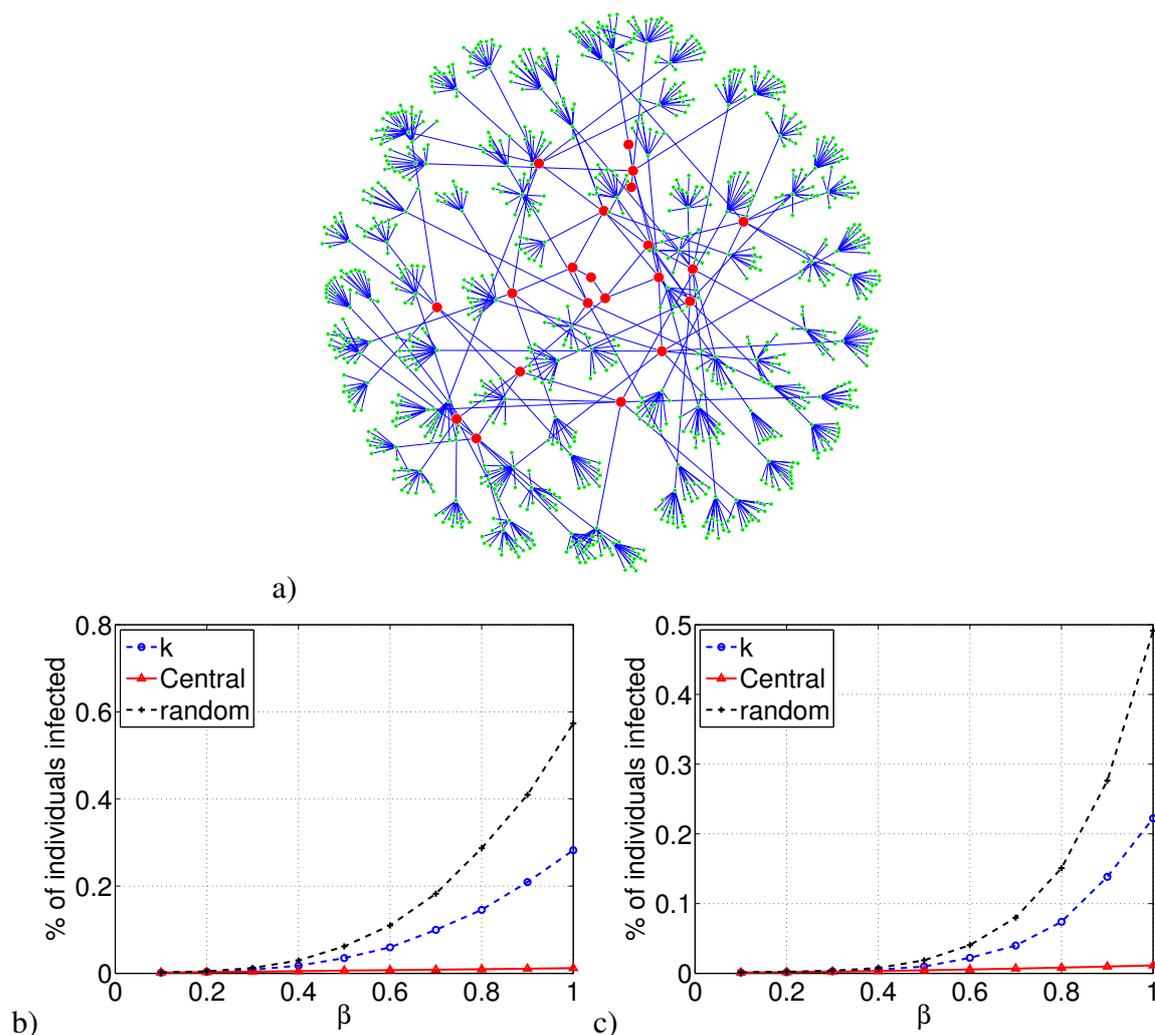


Figura 14 – Avaliação da propagação da doença na rede árvore artificial. (a) de visualização de rede: os nós em forma de vermelho a fragmentação mais central. Esta visualização baseia-se no paradigma de “força-dirigida” conforme implementado por Kamada e Kawai (KAMADA; KAWAI, 1989; SHANNON *et al.*, 2003). Em (b) e (c): porcentagem máxima de indivíduos infectados após os procedimentos de imunização baseada em hubs, cluster mais central e escolha aleatória de nós para $\mu = 0.3$ e $\mu = 0.6$, respectivamente.

Fonte: Elaborada pelo autor.

é pequena. Este é um efeito de falta de centros periféricos em todas as redes do mundo real considerados. Contudo, Kitsak *et al.* (KITSAK *et al.*, 2010) verificaram que os hubs periféricos não são propagadores influentes em redes complexas. A fim de verificar esta propriedade, geramos uma árvore artificial no qual os hubs são colocados na fronteira da rede e comparamos a eficácia da distribuição de tais hubs com a capacidade de os vértices no cluster mais central. A Figura 14 mostra os resultados. A utilização da abordagem de agrupamento de dados permite identificar os vértices no centro da árvore (veja Figura 14(a)). Além disso, as Figuras 14(b) e (c) mostram que a imunização dos nós no cluster mais central é mais eficaz para controlar o processo de propagação do que a vacinação de hubs. Portanto, o aglomerado mais central contém

os propagadores mais influentes na referida rede e o grau de nó não é uma medida adequada para a identificação de propagadores principais neste caso.

4.5 Conclusões

Este capítulo sugeriu um método de agrupamento de dados para determinar os propagadores mais influentes em redes complexas, generalizando o conceito de centralidade rede. A abordagem baseia-se no algoritmo de maximização da esperança e um conjunto de índices de centralidade, evitando a arbitrariedade na escolha de uma medida de centralidade, uma vez que não existe uma métrica que seja suficientemente geral para quantificar o papel central de um nó. Na verdade, a definição de centralidade da rede é geralmente baseada em conceitos diferentes, tais como caminhadas aleatórias ou menores distâncias.

Os vértices com os maiores valores de medidas de centralidade individuais mostraram ser os mais influentes propagadores de doenças. Eles foram identificados automaticamente, considerando-se o algoritmo de maximização de esperança. Comparou-se a capacidade de propagação desses vértices com abordagens que consideram grau e *k-core*. Verificou-se que para todas as redes, vacinar os vértices no cluster mais central é a melhor maneira de diminuir a epidemia. Para as redes sociais e de comunicação, a imunização dos hubs produziu resultados semelhantes em algumas redes, o que sugere que estas redes não mostram uma elevada percentagem de centros periféricos. No entanto, quando analisamos uma árvore artificial, em que os hubs estão na fronteira da rede, verificamos que a nossa abordagem é mais adequado para identificar os propagadores mais influentes.

Estes resultados mostram que o cluster que contém os nós mais centrais em termos de medida multivariada M_i são os propagadores mais influentes do modelo SIR. Portanto, a consideração dos índices de centralidade adicionais permite uma identificação mais precisa dos nós mais centrais e uma generalização do conceito de centralidade. Além disso, o método proposto permite a determinação do número de propagadores influentes automaticamente.

Nosso estudo pode ser estendido para considerar métricas de centralidade adicionais e outros algoritmos de agrupamento. Outras definições do índice M_i pode ser obtida ponderando-se a importância de cada medida. A nossa abordagem pode ser aplicada a outros processos dinâmicos, tais como falhas em cascata e sincronização a fim de identificar os vértices que mais influenciam tais processos.

IDENTIFICAÇÃO DE PROPAGADORES IMPORTANTES USANDO MEDIDAS DE CENTRALIDADE

Nesse capítulo, vamos analisar como propriedades dos vértices influenciam a dinâmica de propagação de epidemias e rumores. Particularmos, vamos analisar como métricas de centralidade estão relacionadas com a capacidade de propagação de cada vértice. Quando a maioria dos agentes centrais podem propagar sua influência a todos de forma mais rápida que o restante dos agentes, acredita-se que estes agentes sejam os propagadores mais importantes. Recentemente, [Poux-Médard, Pastor-Satorras e Castellano \(2020\)](#) mostraram que a medida de grau pode ser importante para identificar os propagadores em epidemias recorrentes, e [Kitsak *et al.* \(2010\)](#) observaram que os principais propagadores de doenças podem ser previstos a partir da análise de decomposição *k-shell*, como [Zhang *et al.* \(2016\)](#) também demonstraram. Já os pesquisadores [Pastor-Satorras e Castellano \(2017\)](#) utilizaram-se do índice de Hirsch para identificar os propagadores mais influentes na rede. Estes últimos localizam-se no núcleo do grafo e não precisam ser os de maior conectividade. Pesquisadores [Silva, Viana e Costa \(2012\)](#) observaram que conectividade e acessibilidade são relacionadas com a difusão mais eficiente de uma epidemia. Ao contrário, [Borge-Holthoefer e Moreno \(2012\)](#) observaram que, no caso dos modelos de rumor, não se pode identificar os disseminadores mais importantes usando as mesmas centralidades.

Mesmo com muitos trabalhos fornecendo evidências para a existência de propagadores influentes de doenças, as conclusões não são generalizadas. Ou seja, não há um consenso comum sobre “centralidade” em grafos. Vimos no capítulo anterior que é possível definir a centralidade usando-se diferentes conceitos, como caminhadas aleatórias ou distância entre vértices ([NEWMAN, 2010](#)). Outro exemplo, onde as medidas de *betweenness* e proximidade consideram apenas a menor distância entre todas as duplas de vértices ([BOCCALETTI *et al.*, 2006](#); [NEWMAN, 2010](#)). Da mesma forma, a decomposição *k-shell* pode deletar vértices que

estão entre comunidades (SEIDMAN, 1983). Deste modo, para suprir a falta de uma definição geral de centralidade do vértice, é importante considerar métricas adicionais. Neste trabalho, estudamos o problema da identificação de propagadores influentes comparando oito medidas de centralidade (KITSACK *et al.*, 2010; BORGE-HOLTHOEFER; MORENO, 2012).

Os resultados e discussões apresentados nesse capítulo são baseados em nosso artigo publicado na revista *Physical Review E* (ARRUDA *et al.*, 2014), onde pode-se encontrar mais detalhes sobre nossa análise.

5.1 Acessibilidade de caminhada aleatória generalizada

A medida de acessibilidade está relacionada com a diversidade no acesso de nós através de caminhadas aleatórias (TRAVENÇOLO; F. Costa, 2008). Para identificar fronteira de grafos foi utilizada a acessibilidade (TRAVENÇOLO; VIANA; COSTA, 2009). Seja $P^{(h)}(i, j)$ a probabilidade para alcançar o vértice j realizando caminhadas aleatórias de comprimento h partida i . A acessibilidade de i para uma distância h fixa é estabelecida pelo exponencial da fórmula de Shannon para entropia (TRAVENÇOLO; F. Costa, 2008), ou seja,

$$\alpha_h(i) = \exp \left(- \sum_j P^{(h)}(i, j) \log P^{(h)}(i, j) \right), \quad (5.1)$$

e $1 \leq \alpha_h(i) \leq N$. O máximo valor será no caso em que todos os agentes são acessados com mesma probabilidade de $1/N$. Note que esta métrica foi definida de forma multi-nível, dependendo do parâmetro h que define a escala da dinâmica (TRAVENÇOLO; VIANA; COSTA, 2009; TRAVENÇOLO; F. Costa, 2008).

A fim de introduzirmos uma generalização da medida de acessibilidade, vamos considerar o conceito de matriz exponencial (BHATIA, 1997). Para a probabilidade de transição toma-se as caminhadas de todos os comprimentos h entre todos pares. Podemos definir a matriz exponencial de P utilizando a matriz de transição P , conforme esta equação:

$$\mathbf{W} = \sum_{k=0}^{\infty} \frac{1}{k!} P^k = e^P. \quad (5.2)$$

Baseando-se em caminhadas aleatórias modificadas, penalizando caminhos maiores pelo inverso do fatorial.

A probabilidade e transição de uma caminhada aleatória generalizada, relacionada com a matriz W é obtida por (ARRUDA *et al.*, 2014)

$$\mathbf{P} = \frac{\mathbf{W}}{e}. \quad (5.3)$$

A expressão generalizada da acessibilidade é definida assim:

$$\alpha(i) = \exp \left(- \sum_j \mathbf{P}(i, j) \log \mathbf{P}(i, j) \right), \quad (5.4)$$

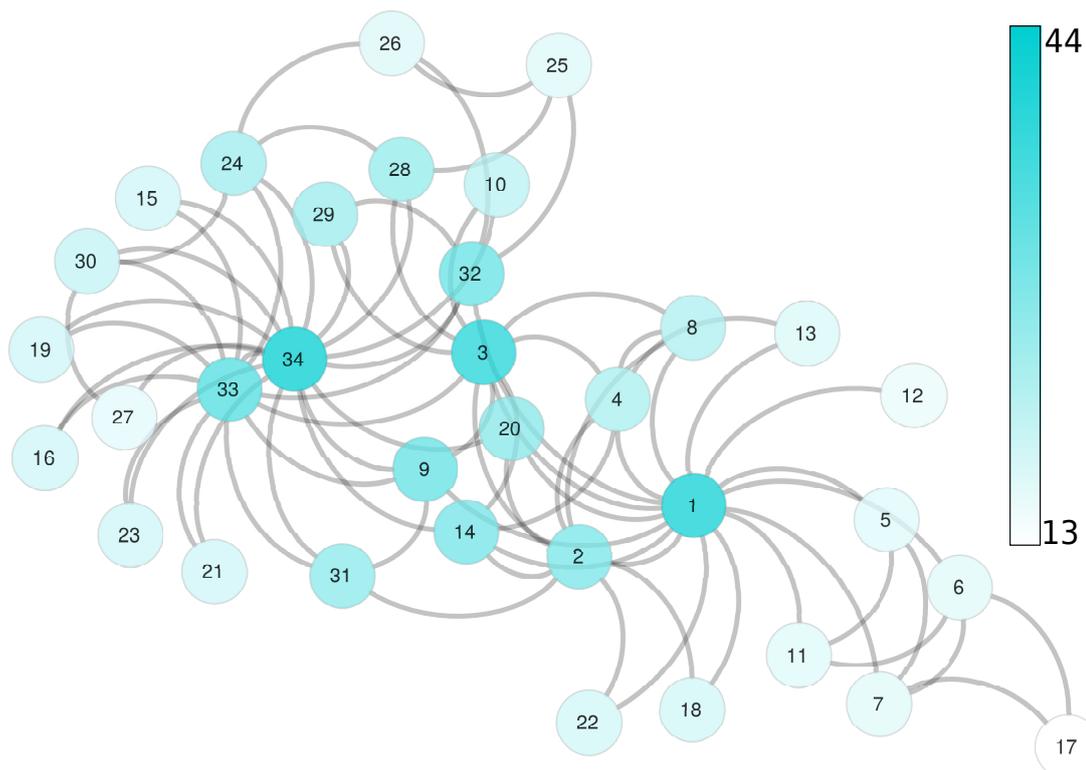


Figura 15 – Gráfico de conceito da medida acessibilidade (valores baesados na Equação 5.4) na rede do clube de Caratê, Zachary (ZACHARY, 1977). Nós centrais apresentam a maior acessibilidade. Figura extraída de (ARRUDA *et al.*, 2014).

onde recebe o nome de acessibilidade generalizada por caminhada aleatória. A Figura 15 ilustrada o cálculo dessa medida.

Essa matriz exponencial foi também considerada quando se definiu a medida comunicabilidade (ESTRADA; HATANO; BENZI, 2011; ESTRADA; HATANO, 2008). A diferença entre essa medida e a acessibilidade generalizada é que a acessibilidade é baseada no conceito de diversidade (HILL, 1973; JOST, 2006)

Para iniciarmos nossa análise sobre a acessibilidade generalizada, vamos calcular o valor exato dessa medida para alguns grafos regulares.

5.1.1 Acessibilidade de grafo em estrela

Utilizando um grafo no formato estrela, a probabilidade de transicionar entre o vértice central i e outro qualquer dos k nós folhas é dada pela equação:

$$\mathbf{P}(i, j) = \frac{1}{ek} \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} = \frac{1}{ek} \sinh(1), \quad i \neq j \quad (5.5)$$

já entre as folhas e o vértice central i ,

$$\mathbf{P}(j, i) = \frac{\sinh(1)}{e}. \quad (5.6)$$

Assim,

$$\mathbf{P}(i, i) = \frac{\cosh(1)}{e}. \quad (5.7)$$

A probabilidade de transição entre j e l é:

$$\mathbf{P}(j, l) = \frac{1}{ek}(\cosh(1) - 1), \quad (5.8)$$

e quando $l = j$,

$$\mathbf{P}(j, j) = \frac{1}{e} + \frac{1}{ek}(\cosh(1) - 1). \quad (5.9)$$

Portanto, a forma geral da matriz exponencial em um gráfico de estrela é:

$$\mathbf{P} = \frac{1}{e} \begin{bmatrix} \cosh(1) & \frac{1}{k} \sinh(1) & \cdots & \cdots & \frac{1}{k} \sinh(1) \\ \sinh(1) & 1 + \frac{1}{k}(\cosh(1) - 1) & \frac{1}{k}(\cosh(1) - 1) & \cdots & \frac{1}{k}(\cosh(1) - 1) \\ \vdots & \frac{1}{k}(\cosh(1) - 1) & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \frac{1}{k}(\cosh(1) - 1) \\ \sinh(1) & \frac{1}{k}(\cosh(1) - 1) & \cdots & \frac{1}{k}(\cosh(1) - 1) & 1 + \frac{1}{k}(\cosh(1) - 1) \end{bmatrix}. \quad (5.10)$$

Assim, quando $k = N - 1$, a medida de acessibilidade de i , o hub, será

$$\alpha(i) = \exp \left\{ -x \log(x) - y \log \left(\frac{y}{N-1} \right) \right\}, \quad (5.11)$$

e $x = \frac{\cosh(1)}{e}$ e $y = \frac{\sinh(1)}{e}$. Tomando qualquer folha j ligada a i ,

$$\alpha(j) = \exp \{ -x \log(x) + (N-2)y \log(y) + (1/e + y) \log(1/e + y) \}, \quad (5.12)$$

e $x = \frac{\sinh(1)}{e}$, $y = \frac{(\cosh(1)-1)}{e(N-1)}$.

Na Figura 16, temos os resultados obtidos para a acessibilidade para diferentes redes e configurações. Como vemos, a Equação 5.11 pode ser utilizada para prever a acessibilidade de hubs em redes SF. Porém, esta aproximação de gráfico estrela não é precisa para redes aleatórias ER.

5.1.1.1 Decompondo matriz

Pela decomposição da matriz P , podemos obter valores exatos de acessibilidade de grafos estrela. A matriz exponencial W pode ser obtida da seguinte forma,

$$\mathbf{W} = e^P = \mathcal{V} \mathcal{D} \mathcal{V}^{-1}. \quad (5.13)$$

onde \mathcal{V} é a matriz cuja as linhas verticais são autovetores da matriz \mathbf{W} , e a diagonal de \mathcal{D} apresenta a exponencial de cada autovalor de P ,

$$(\mathbf{P} - \lambda \mathbf{I}) v = 0 \quad (5.14)$$

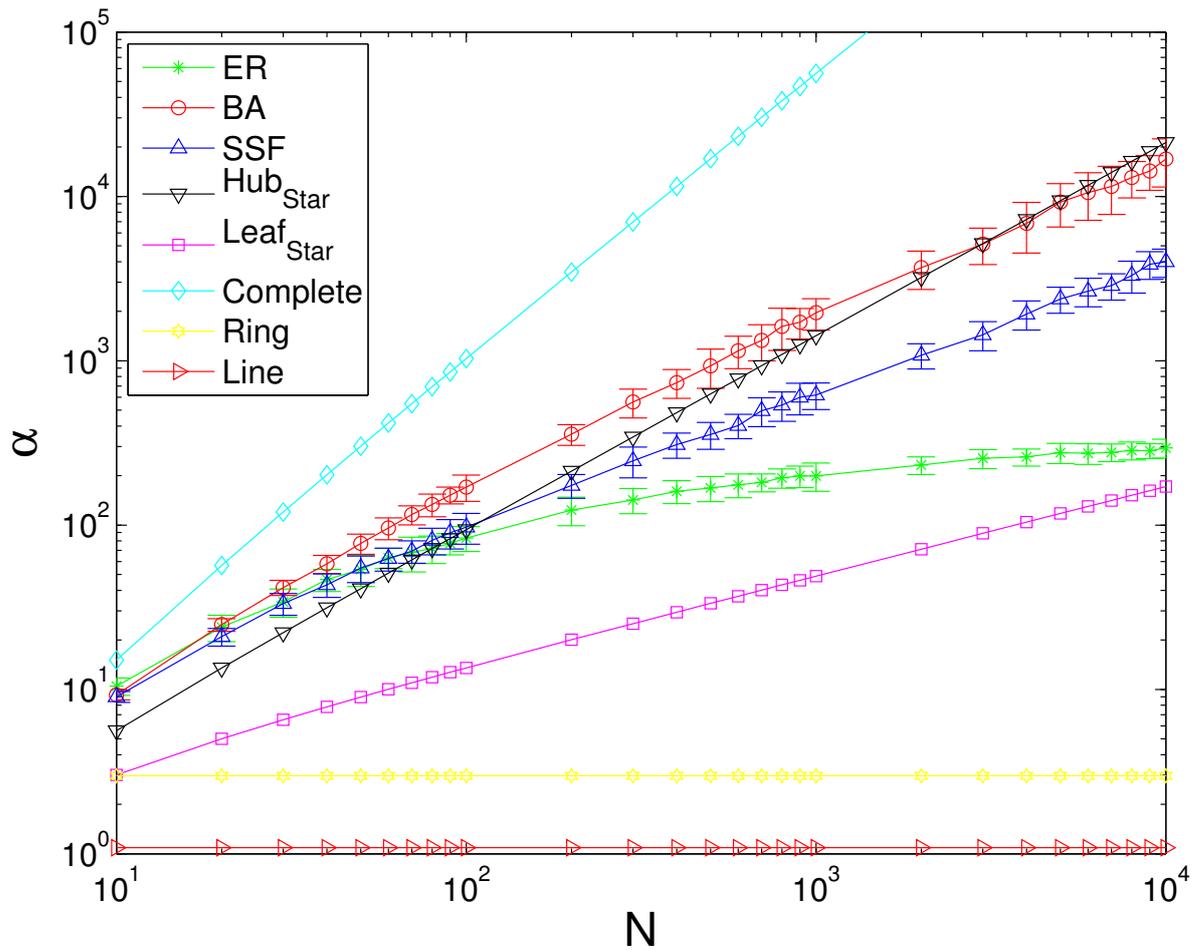


Figura 16 – Acessibilidade calculada para a estrela (a partir das Eqs. 5.12 e 5.11), grafos completamente conectados, linha e anel em comparação com o valor máximo obtido nos grafos (ER), redes (BA), e redes livre de escala espaciais (S S F). N é o número de nós. Cada ponto é uma média de +50 redes de $\langle k \rangle \approx 4$. Figura extraída de (ARRUDA *et al.*, 2014).

onde P é a matriz de transição, λ é seu autovalor e v é o autovetor associado.

Assim, no caso de grafo estrela, a matriz P de transição é esparsa e tem polinômio bem característico, $\det(P - \lambda I) = 0$, que pode ser calculado por Laplace, conforme abaixo,

$$\begin{aligned} \det(P - \lambda I) &= (-\lambda)^N - (-\lambda)^{(N-2)} = \\ &= (-\lambda)^{(N-2)} \left((-\lambda)^2 - 1 \right) = 0, \end{aligned} \quad (5.15)$$

onde as soluções são $\lambda_1 = -1$, $\lambda_2 = 1$ e $\lambda_i = 0, \forall i = 2, 3, \dots, N$. Assim, considerando-se esta definição como um problema de autovalores e autovetores, será possível encontrar as equações dos autovetores. Seja $\lambda_1 = -1$,

$$\begin{cases} v_{11} = -\frac{1}{N-1} \sum_{j=2}^N v_{1j} \\ v_{1j} = -v_{11} \quad j = 2, 3, \dots, N; \end{cases} \quad (5.16)$$

e v_{pj} como o j -ésimo membro do autovetor v_p que se associa ao autovalor λ_p . Para $\lambda = 1$,

$$\begin{cases} v_{21} = \frac{1}{N-1} \sum_{j=2}^N v_{2j} \\ v_{2j} = v_{21} \quad j = 2, 3, \dots, N; \end{cases} \quad (5.17)$$

finalmente, para $\lambda_p = 0$ onde $p = 3, \dots, N$, que tem multiplicidade $(N-2)$,

$$\begin{cases} 0 = \frac{1}{N-1} \sum_{j=2}^N v_{pj} \\ v_{p1} = 0 \end{cases} \quad (5.18)$$

que produz a matriz

$$\mathcal{V} = \begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & -1 & \cdots & \cdots & -1 \\ \vdots & \vdots & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 1 & 1 & 0 & \cdots & 0 & 1 \end{bmatrix}, \quad (5.19)$$

e a inversa é

$$\mathcal{V}^{-1} = \begin{bmatrix} \frac{-1}{2} & \frac{1}{2(N-1)} & \frac{1}{2(N-1)} & \cdots & \frac{1}{2(N-1)} \\ \frac{1}{2} & \frac{1}{2(N-1)} & \frac{1}{2(N-1)} & \cdots & \frac{1}{2(N-1)} \\ 0 & \frac{-1}{N-1} & \frac{N-2}{N-1} & & \frac{-1}{N-1} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \frac{-1}{N-1} & \cdots & & \frac{N-2}{N-1} \end{bmatrix}. \quad (5.20)$$

Usamos vetores não-unitários para construir as matrizes \mathcal{V} , visto que \mathcal{D} também é multiplicado por \mathcal{V}^{-1} e as normas não-unitárias são compensadas. Substituindo nas matrizes 5.19 and 5.20 da Eq. 5.13, de depois de aplicar álgebra, chegamos na Eq. 5.10. A acessibilidade das folhas e dos *hubs* é calculada através das Equações 5.12 e 5.11, respectivamente.

5.1.2 Acessibilidade de grafo em anel

A acessibilidade usando caminhada aleatória da forma generalizada pode ser calculada de forma exata para anéis, um caso especial de grafos regulares, para $K = 2$. A matriz que tem as probabilidades de transicionar apresenta a forma

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & \ddots & \ddots & & & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & & \ddots & \ddots & \frac{1}{2} \\ \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} & 0 \end{bmatrix}. \quad (5.21)$$

Tal matriz tem um espectro bem conhecido e é largamente usada em métodos de diferenças finitas (LEVEQUE, 2007). como exposto por LeVeque (2007), os autovalores de P são

$$\lambda_p = \frac{1}{2} \left(\exp \left(\frac{2\pi ip}{N} \right) + \exp \left(\frac{-2\pi ip}{N} \right) \right) = \cos \left(\frac{2\pi p}{N} \right), \quad (5.22)$$

onde $i = \sqrt{-1}$ e os elementos associados do autovetor podem ser expressos como

$$u_{pj} = \frac{\exp \left(\frac{2\pi ipj}{N} \right)}{\sqrt{N}}, \quad (5.23)$$

e \sqrt{N} é somente um fator para normalizar. Os autovetores diagonalizam a matriz P da forma $P = U\Lambda U^H$, e Λ é a matriz que em sua diagonal tem autovalores de P (Eq. 5.22), U é a matriz em que suas colunas são os autovetores de P , e U^H é o transposto conjugado. Podemos escrever a expressão fechada para \mathbf{P} conforme

$$\mathbf{P}(j, k) = \frac{1}{e} \sum_p \exp(\lambda_p) u_{pj} u_{pk}^*, \quad (5.24)$$

onde u_{pj}^* é o transposto conjugado de u_{pj} . Note que utilizamos o domínio complexo para resolver este problema, porém, a solução fica no domínio real. Utilizando as Equações 5.22 e 5.23, nós obtemos

$$\mathbf{P}(j, k) = \frac{1}{eN} \sum_{p=1}^N \exp \left(\cos \left(\frac{2\pi p}{N} \right) \right) \exp \left(\frac{2\pi ip(j-k)}{N} \right), \quad (5.25)$$

que é uma forma fechada para a avaliação de \mathbf{P} grafos em forma de anel. Além disso, podemos usar algumas propriedades espectrais do grafo para separar o primeiro autovalor do somatório, conforme

$$\begin{aligned} \mathbf{P}(j, k) &= \frac{1}{e} \left(\frac{k_j}{2M} \right) \exp(1) + \\ & \frac{1}{N} \sum_{p=2}^N \exp \left(\cos \left(\frac{2\pi p}{N} \right) + \left(\frac{2\pi ip(j-k)}{N} \right) \right). \end{aligned} \quad (5.26)$$

A Figura 16 mostra os resultados para diversos grafos regulares, incluindo o grafo em anel e o grafo em linha.

5.1.3 Acessibilidade em grafo completo

Podemos calcular a acessibilidade para um grafo todo completo, onde todos vértices estão interligados. A probabilidade de transicionar entre qualquer par é $P(i, j) = \frac{1}{N-1}$ e a matriz exponencial é dada por

$$\begin{aligned} \mathbf{P}(i, j) &= \frac{1}{eN} \sum_{n=0}^{\infty} \frac{(N-1)^n + (-1)^n (N-1)}{(N-1)^n n!} = \\ & \frac{\exp(1) + (N-1) \exp \left(\frac{-1}{N-1} \right)}{eN}, \quad i \neq j. \end{aligned} \quad (5.27)$$

Já a diagonal principal \mathbf{P} leva em conta os caminhos começando e terminando no mesmo ponto, como a seguir:

$$\mathbf{P}(i, i) = \frac{1}{eN} \sum_{n=0}^{\infty} \frac{(N-1)^n + (-1)^{n+1}}{(N-1)^n n!} = \frac{\exp(1) - \exp\left(\frac{-1}{N-1}\right)}{eN}. \quad (5.28)$$

Portanto, temos a forma geral da matriz exponencial:

$$\mathbf{P} = \frac{1}{e} \begin{bmatrix} \frac{\exp(1) + (N-1)\exp\left(\frac{-1}{N-1}\right)}{N} & \frac{\exp(1) - \exp\left(\frac{-1}{N-1}\right)}{N} & \cdots & \frac{\exp(1) - \exp\left(\frac{-1}{N-1}\right)}{N} \\ \frac{\exp(1) - \exp\left(\frac{-1}{N-1}\right)}{N} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\exp(1) - \exp\left(\frac{-1}{N-1}\right)}{N} & \cdots & \cdots & \frac{\exp(1) + (N-1)\exp\left(\frac{-1}{N-1}\right)}{N} \end{bmatrix}. \quad (5.29)$$

A acessibilidade de cada vértice é

$$\begin{aligned} \alpha(i) &= \exp\{-\mathbf{P}(i, i) \log(\mathbf{P}(i, i)) + \\ &\quad -(N-1)\mathbf{P}(i, j) \log(\mathbf{P}(i, j))\} = \\ &\quad e \left(a^{-(N-1)a/e} b^{-b/e} \right), \end{aligned} \quad (5.30)$$

onde

$$a = \frac{1}{eN} \left(\exp(1) - \exp\left(\frac{-1}{N-1}\right) \right) \quad (5.31)$$

e

$$b = \frac{1}{eN} \left(\exp(1) + (N-1)\exp\left(\frac{-1}{N-1}\right) \right). \quad (5.32)$$

No grafo completo, todos os vértices apresentam o mesmo valor de acessibilidade, sendo a acessibilidade máxima possível dentre todas as redes com N nós. Na Figura 16 vemos a variação da acessibilidade para grafos completos em função do tamanho do grafo.

Valores exatos da acessibilidade para grafos completos, podem ser obtido através da decomposição da matriz, como mostramos antes. Assim, teremos o sistema:

$$\begin{bmatrix} -\lambda & \frac{1}{N-1} & \cdots & \frac{1}{N-1} \\ \frac{1}{N-1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{N-1} & \cdots & \cdots & -\lambda \end{bmatrix} v = 0, \quad (5.33)$$

que produz

$$\sum_{i, i \neq j}^N v_{pi} \left(\frac{1}{N-1} \right) - \lambda_p v_{pj} = 0, \quad \forall i, j; i \neq j, \quad (5.34)$$

onde v_{pj} é o j -ésimo elemento do autovetor v_p associado com o autovalor λ_p .

Autovalores de P de um grafo completo são o espectro da matriz A multiplicados por $\frac{1}{N-1}$, i.e., $\lambda_1 = 1, \lambda_2 = \lambda_3 = \dots = \lambda_N = \frac{1}{N-1}$ (MIEGHEM, 2011). Assim, se $\lambda_1 = 1$ teremos

$$(N-1)v_{1j} = \sum_{i, i \neq j}^N v_{1i}, \quad \forall i, j; i \neq j. \quad (5.35)$$

A solução fica $v_{1i} = v_{1j}$. Por outro lado, para $\lambda_p = \frac{-1}{N-1}$, onde $p = 2, 3, \dots, N$,

$$\sum_{i, i \neq j}^N v_{pi} \left(\frac{1}{N-1} \right) + \frac{1}{N-1} v_{pj} = 0, \quad (5.36)$$

isto é,

$$\left(\frac{1}{N-1} \right) \sum_i^N v_{pi} = 0, \quad \forall i, j; \quad i \neq j \quad (5.37)$$

A respectiva solução é $\sum_i^N v_{pi} = 0$. As duas soluções não são únicas, já a Equação 5.13 tem solução única. Tomemos

$$\mathcal{V} = \begin{bmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & 0 & \cdots & & 1 \end{bmatrix}, \quad (5.38)$$

e sua inversa

$$\mathcal{V}^{-1} = \begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{-1}{N} & \frac{(N-1)}{N} & \frac{-1}{N} & \cdots & \frac{-1}{N} \\ \vdots & \frac{-1}{N} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \frac{-1}{N} & \frac{-1}{N} & \cdots & & \frac{(N-1)}{N} \end{bmatrix}. \quad (5.39)$$

Para substituir as matrizes \mathcal{V} e \mathcal{V}^{-1} na Equação 5.13 e utilizando a informação sobre os autovalores, obtém-se a matriz da Eq. 5.29 e a de volta a expressão da acessibilidade (Equação 5.30).

5.2 Propagação de epidemias e rumores

Vamos analisar o comportamento entre as métricas de centralidade e a capacidade de propagar de cada vértice. Nós adotamos que a dinâmica de difusão inicia em um único nó, e os demais nós ficam no estado de suscetíveis (ou ignorantes, para rumor). No modelo de epidemia SIR, a capacidade de propagar de cada nó é quantificada pela prevalência do processo de propagação. A capacidade de difundir i é o percentual de vértices recuperados no final da dinâmica, uma vez que o processo inicia em i , ou seja, $H(i) = r(t \rightarrow \infty)$. Desta forma, a capacidade de difundir de um vértice i em uma dinâmica do tipo rumor é quantificada pelo percentual de *stiflers* no final da dinâmica, sendo que a propagação inicia em i , ou seja, $M(i) = s(t \rightarrow \infty)$.

5.3 Base de dados

Foram realizadas simulações numéricas de processos epidêmicos e de rumores em redes reais e artificiais. A Tabela 3 mostra as propriedades topológicas de rede rodoviária de alguns

paíse e também redes artificiais.

Tabela 3 – Medidas topológicas das redes da base de dados.

	Network	N	$\langle k \rangle$	$\langle cc_i \rangle$	$\langle B_i \rangle$	$\langle C_i \rangle$	$\langle r_i \rangle$	$\langle \pi \rangle$	$\langle \alpha \rangle$	$\langle x_i \rangle$	$\langle k_c \rangle$
Spatial	Japão	2130	3.792	0.24	3.731×10^4	0.03	4.290	4.695×10^{-4}	6.950	2.892×10^{-3}	2.523
	Inglatera	4460	3.415	0.14	8.163×10^4	0.03	3.557	2.242×10^{-4}	6.650	1.401×10^{-3}	2.062
	Estados Unidos	6443	3.098	0.09	1.605×10^5	0.02	3.302	1.552×10^{-4}	6.178	9.328×10^{-4}	2.038
	Alemanha	3555	3.068	0.08	5.944×10^4	0.03	3.173	2.813×10^{-4}	6.243	2.668×10^{-3}	1.988
	Spatial SF	5000	3.998	0.04	1.226×10^4	0.17	9.291	2.000×10^{-4}	9.793	6.001×10^{-3}	2.000
	Wax man	4883	4.078	0.14	4.598×10^4	0.05	4.863	2.048×10^{-4}	8.071	1.433×10^{-3}	2.570
Non-spatial	advogado	5054	15.58	0.25	5.748×10^3	0.31	99.62	1.979×10^{-4}	28.92	6.819×10^{-3}	8.137
	e-mail	1133	9.622	0.22	1.475×10^3	0.28	17.90	8.826×10^{-4}	17.88	1.764×10^{-2}	5.349
	Political blogs	1222	27.36	0.32	1.061×10^3	0.37	100.1	8.183×10^{-4}	33.08	1.681×10^{-2}	14.82
	Google+	23613	3.319	0.17	3.580×10^4	0.25	727.0	4.235×10^{-5}	15.13	2.301×10^{-3}	1.669
	BA	10000	3.999	0.01	2.005×10^4	0.20	17.06	1.000×10^{-4}	10.57	3.108×10^{-3}	2.000

As redes de rodovias foram extraídas a partir de mapas disponíveis no formato digital portátil (*pdf*) no *website* das Nações Unidas ¹. Na etapa inicial, pré processamos os mapas para retirar informações que não sejam relevantes e deixar apenas as rodovias importantes. Na sequência, realizamos o processo de esquelizar a imagem, a fim de extrair o esqueleto (COSTA; Cesar Jr, 2000). A identificação de cada nó foi realizada pela aplicação de um filtro de convolução acerto-erro de 8-conectados (DOUGHERTY, 1992). Por fim, propaga-se rótulos a partir de cada vértice. Assim que dois rótulos se encontram uma ligação é formada. Consideramos extrair redes dos mapas dos países: Estados Unidos, Alemanha, Japão, Inglaterra.

Já o grafos sociais considerados aqui são: (a) rede de mensagens obtida a partir de emails trocadas pelos usuários da *Universidade Rovira e Virgili* (GUIMERA *et al.*, 2003); (b) rede de *blogs* de políticos, formada por hiperlinks de *blogs* durante o período de meses próximos a eleição de presidente americano do ano 2004 (ADAMIC; GLANCE, 2005); (c) rede Advogado, uma comunidade online dedicada a desenvolver software livre que iniciou-se no ano de 1999 (ADVOGATO..., 2014; MASSA; SALVETTI; TOMASONI, 2009) e (d) rede social Google+, composta por usuários conectados conforme seus círculos de amigos (MCAULEY; LESKOVEC, 2012; GOOGLE+..., 2014).

5.4 Análise dos parâmetros

O potencial de propagar de um nó vai depender não apenas de suas características estruturais, mas também dos parâmetros usados nos modelos de epidemia. Assim, calculamos as relações entre capacidade de propagar e : a acessibilidade, α , a conectividade, o k , o k -shell, k_c . Cobrimos todos os valores para os parâmetros ($\beta \times \mu$ para a difusão de epidemia e $\lambda \times \delta$ para a difusão de rumor). A Figura 17 demonstra os resultados das redes rodoviárias da Alemanha e dos

¹ <http://www.un.org>

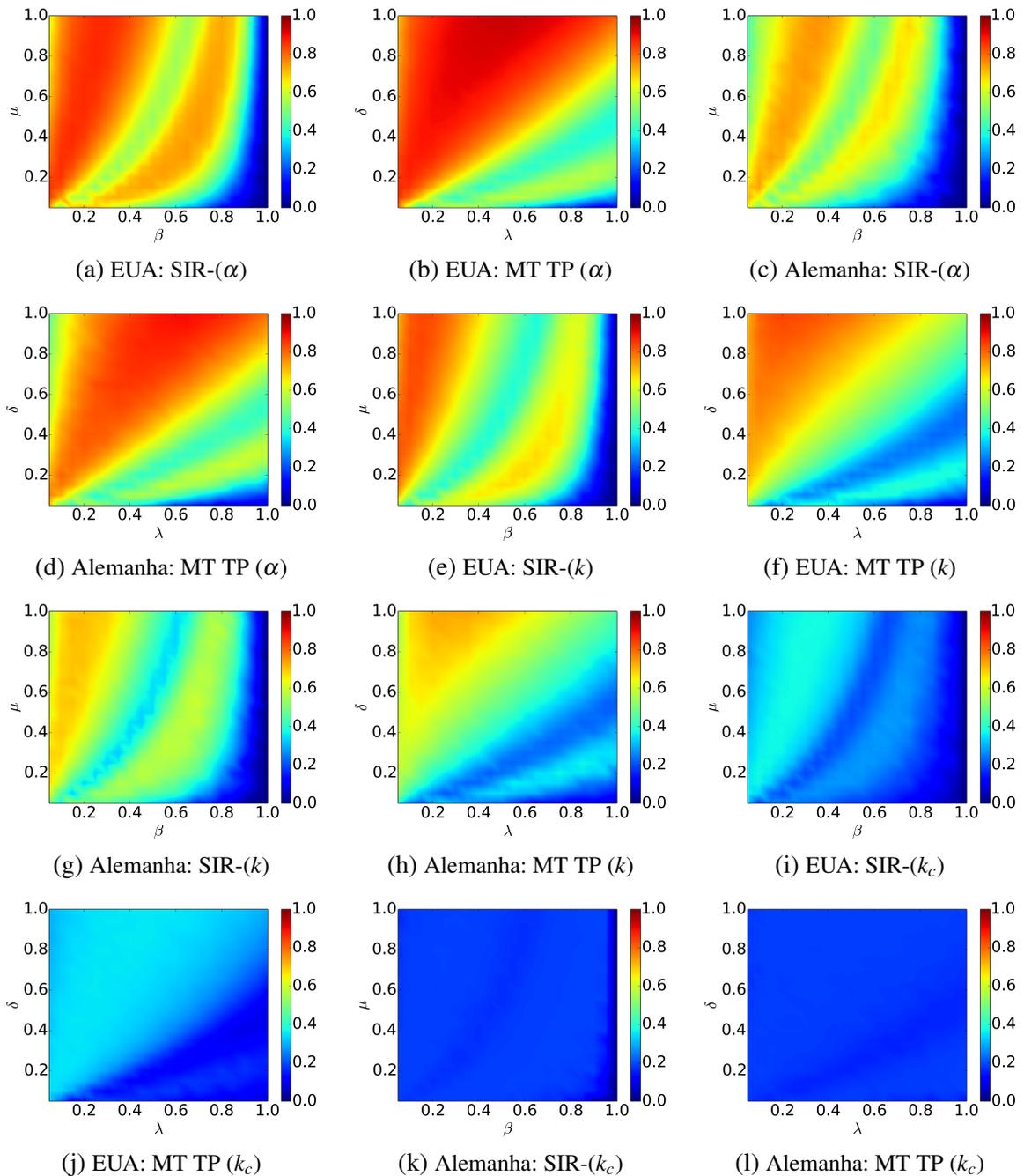


Figura 17 – Relação entre propagação e conectividade, k , acessibilidade, α e k -core, k_c , tomando todos valores dos parâmetros ($\beta \times \mu$ para a difusão de epidemia (SIR) e $\lambda \times \delta$ para a difusão de rumor (MT TP)), em redes de rodovias da Alemanha e dos EUA. Figura extraída de (ARRUDA *et al.*, 2014).

Estados Unidos. Observem que a acessibilidade é a métrica mais relacionada com o potencial de propagar nos dois casos. Porém, a métrica k -shell mostra menor relação para todos os parâmetros.

Com relação ao grau, vemos que essa medida é a mais correlacionado com a capacidade de propagação no modelo SIR para pequenos valores de β . O que era esperado, já que propagar seleciona apenas vizinhos próximos do vértice infectado e, portanto, os vértices de maior

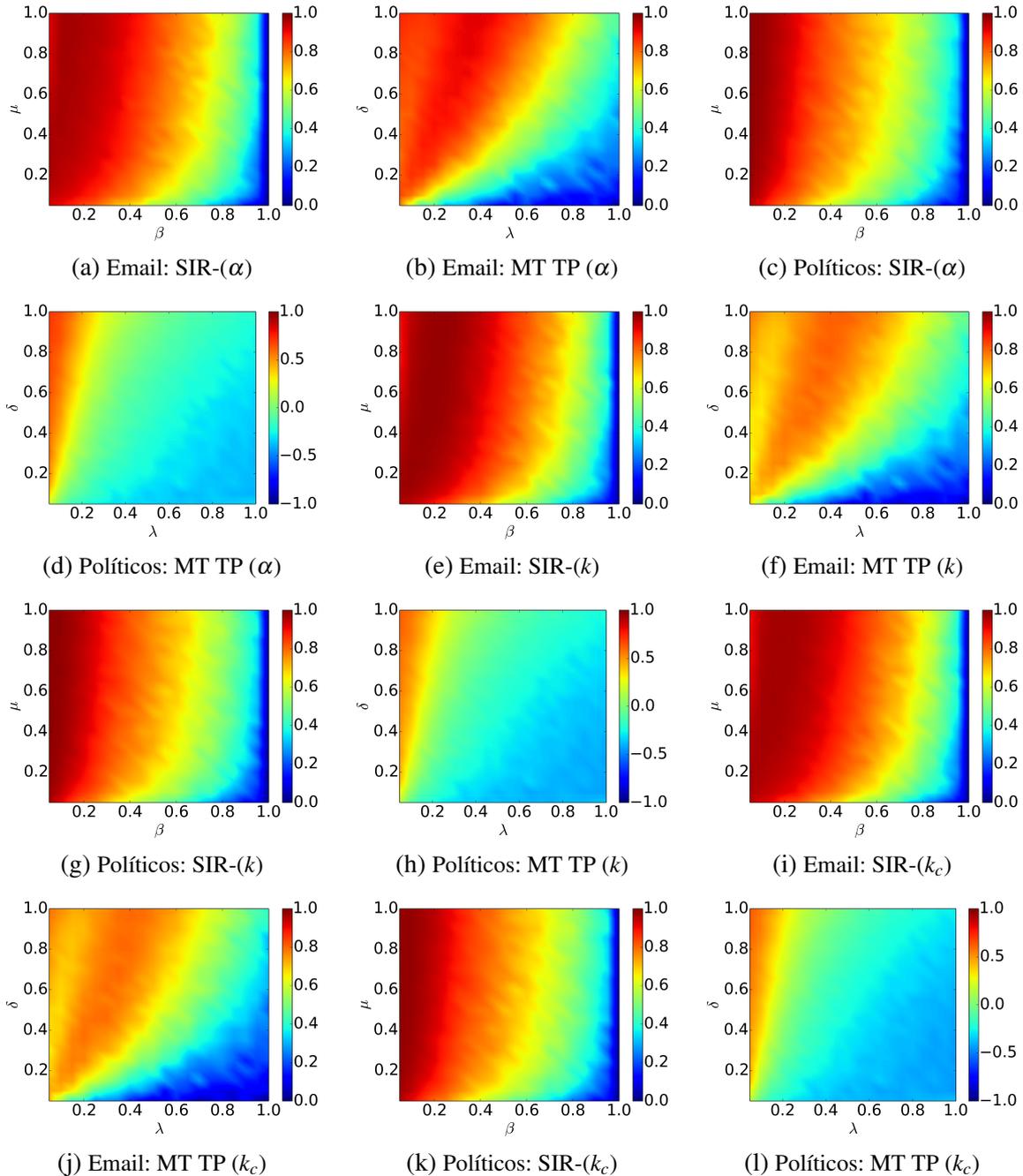


Figura 18 – Relação entre propagação e conectividade, k , acessibilidade, α e k -core, k_c , tomando todos os valores dos parâmetros ($\beta \times \mu$ para a difusão de epidemia (SIR) e $\lambda \times \delta$ para a difusão de rumor (MT TP), para as redes sociais : e-mails e blogs de políticos. Figura extraída de (ARRUDA *et al.*, 2014).

conectividade são os mais importantes na dinâmica de propagação. Observamos que a relação cai para β se aproximando de um, pois a propagação vai selecionar a maioria dos vértices, não dependendo do vértice inicial.

Já na difusão de rumor, a relação entre a medida de acessibilidade e o percentual final de *stiflers* é o mais alta quando $\delta > \lambda$. Esta relação foi também observada para a conectividade,

porém as relações são menores do que as observadas na métrica da acessibilidade.

A Figura 18 expõe os resultados desta análise das redes de mensagens e blogs de políticos. As medidas conectividade, *k-shell* e a medida de acessibilidade geram relações parecidas. Isso é devido a estes grafos não terem hubs na periferia. As maiores relações foram para valores pequenos de λ e β , mostrando que as epidemias e rumores não se difundem após contar os vizinhos mais próximos da origem, portanto, os vértices com maior conectividade são os mais importantes na propagação. Esses vértices também têm os valores maiores das medidas acessibilidade e *k-shell*. Para λ e β próximos de 1, a relação é próxima de zero, como esperado, pois a maioria dos vértices tornam-se *stiflers* ou recuperados não dependendo do propagador que inicia.

Nas próximas seções consideramos (a) $\mu = 1.0, \beta = 0.8$ e (b) $\mu = 1.0, \beta = 0.3$ para a propagação de epidemia e (a) $\delta = 1.0, \lambda = 0.8$, (b) $\delta = 0.3, \lambda = 0.8$, (c) $\delta = 1.0, \lambda = 0.3$ e (d) $\delta = 0.3, \lambda = 0.3$ para dinâmica de rumores. Observe que alegações anteriores (BORGEHOLTHOEFER; MORENO, 2012; KITSACK *et al.*, 2010) usaram $\delta = 1$ e $\mu = 1$ para a propagação de epidemia e rumor, respectivamente. Aqui, consideramos também os casos em que a taxa de propagação é maior do que a taxa de transformação para *stifler*, e uma situação em que ambos têm os mesmos valores. Estes indicam os casos em que os processos são dependentes da organização de rede, ou seja, as relações são maiores que 0, como observamos nas Figuras 18 e 17.

5.5 Redes espaciais

Vemos na Tabela 3 os valores médios para as redes com rodovias e redes geradas pelo modelos de Waxman e o espacial livre de escala. Redes espaciais são esparsas, têm caminhos longos e coeficientes de aglomeração não nulos.

Fizemos simulações dos modelos de epidemia SIR e rumor MT a fim de analisar as relações entre métricas de centralidade dos vértices e o potencial de propagar. Essas correlações foram determinados pelo coeficiente de correlação de Spearman, definido pelo de Pearson para os valores ordenados, ou seja, a ordem das observações (WOLFE; HOLLANDER, 1973). O motivo desta escolha é que o coeficiente proposto por Spearman mede as correlações monótonas, já a relação proposta por Pearson mede apenas correlações lineares.

As Figuras 20 e 19 contêm os resultados de propagação da dinâmica de rumores e epidemia na rede de rodovias dos Estados Unidos da América, respectivamente. A relação mais importante corresponde a conectividade, já para outras medidas, as relações são fracas mas positiva, não zeradas. Contudo, o coeficiente de clustering mostra uma relação negativa pois, quanto mais ligado um nó estiver, menor será o coeficiente de clustering. Por outro lado, as Figuras 21 e 22 mostram que as correlações entre a acessibilidade generalizada e a capacidade de propagação de rumores e epidemia são quase lineares e positivas para todas as redes de rodovias analisadas.

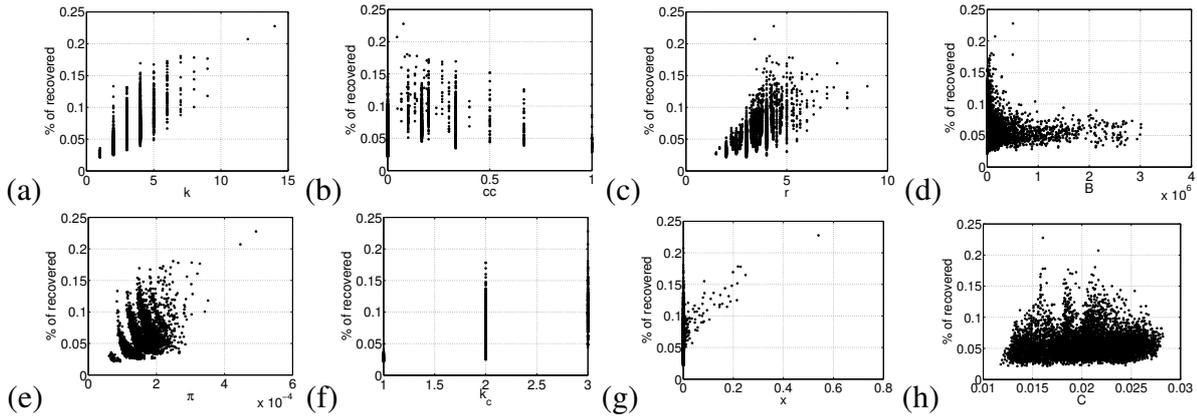


Figura 19 – A porcentagem dos agentes recuperados no modelo de propagação SIR ($\mu = 1.0, \beta = 0.3$) conforme as medidas locais das rede de rodovia dos EUA: (a) conectividade; (b) coeficiente de clustering; (c) grau médio dos vizinhos; (d) *betweenness*; (e) Page Rank; (f) k-shell; (g) autovetores e (h) proximidade. Figura extraída de (ARRUDA *et al.*, 2014).

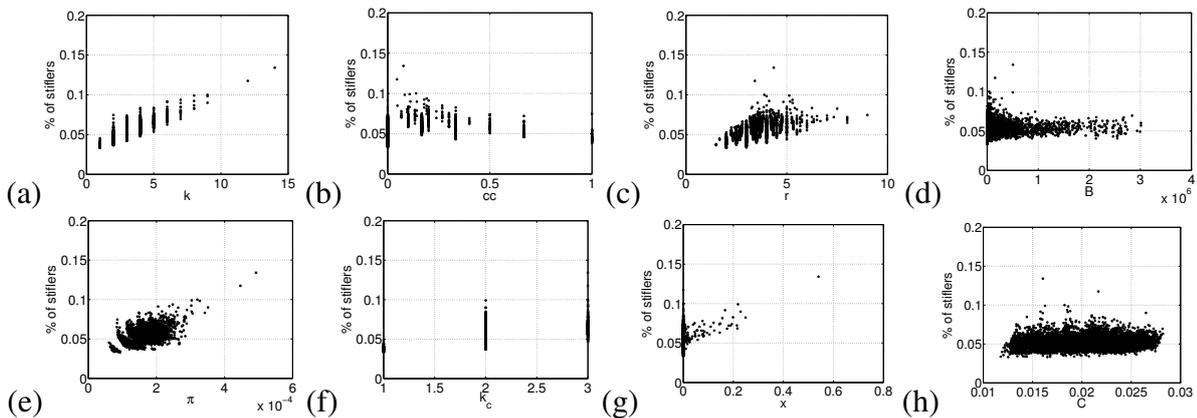


Figura 20 – A porcentagem de *stifiers* no modelo de rumor MT ($\delta = 1.0, \lambda = 0.3$) conforme as medidas locais para a rede de rodovias dos EUA: (a) a conectividade; (b) coeficiente de clustering; (c) grau médio dos vizinhos; (d) *betweenness*; (e) Page Rank; (f) k-shell; (g) autovetores e (h) proximidade. Figura extraída de (ARRUDA *et al.*, 2014).

Adicionalmente, vemos na tabela 4 os dois processos de difusão, as maiores relações entre centralidade e propagação da doença ou rumor corresponde ao caso em que a acessibilidade generalizada é muitas vezes maior que 0,7. Notemos, que a medida *k-shell* gera valores de relação baixa, inverso ao que foi observado em (KITSACK *et al.*, 2010), que considerou grafos não espaciais. Porém, este comportamento está de acordo com Borge-Holthoefler e Moreno (2012), para o processo de rumor. A conectividade do vértice está altamente relacionada com percentual final de vértices recuperados, mas menos relacionada se observarmos os resultados correspondentes ao percentual final de *stifiers*, mais para o caso do modelo de rumor MT utilizando uma configuração CP, assim como encontrado em (BORGE-HOLTHOEFER; MORENO, 2012). Além disso, as medidas de PageRank, *closeness centrality* e *betweenness centrality* não mostram correlações significativas com capacidades propagar doenças e rumor – exceto

quando o parâmetro δ em modelos de rumor é pequeno. Também notamos que a centralidade de autovetor mostra alta correlação apenas para o modelo de rede espacial livre de escala.

5.5.1 Redes de rodovias

Na Figura 23 mostramos os resultados usando a acessibilidade generalizada para redes de estradas do Japão, Inglaterra, Estados Unidos e Alemanha. Manchester, Liverpool e Londres têm os valores maiores da medida de acessibilidade para a Inglaterra, enquanto que no caso dos EUA, as cidades com o maior valor de acessibilidade são Chicago, Nova York, Houston, Dallas. Já no Japão vemos que os propagadores mais importantes são as cidades de Hiroshima, Osaka e Nagoya. Tóquio tem muitas conexões, mas baixa capacidade de propagar, entre estas cidades, já que é hub da periferia. Já na Alemanha, as cidades de Munique, Dusseldorf e Berlim têm os maiores valores de acessibilidade. Observe que os vértices na fronteira entre países têm os menores valores da medida acessibilidade. Logo, esta medida de acessibilidade pode ser usada para identificar a fronteira dos grafos, como observado originalmente na definição de acessibilidade em redes (TRAVENÇOLO; VIANA; COSTA, 2009).

A Figura 24 mostra a distribuição de probabilidade da medida acessibilidade. Em todos casos, a distribuição não é simétrica, mostrando uma cauda estendida para valores maiores da medida acessibilidade. É importante observar que a Inglaterra e a Alemanha apresentam a menor variação da medida acessibilidade.

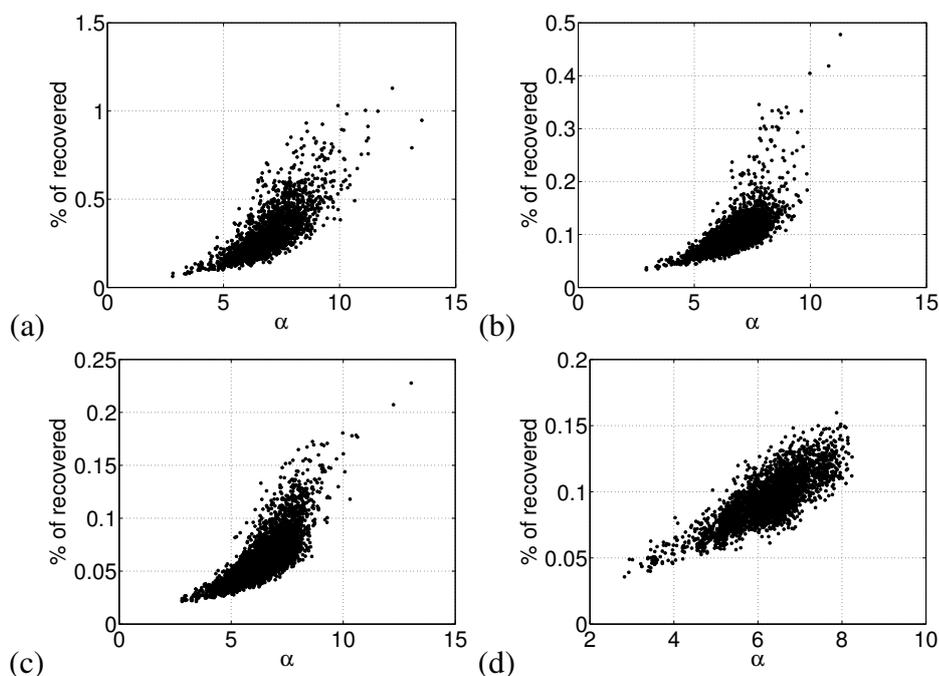


Figura 21 – A porcentagem dos agentes recuperados no modelo de propagação SIR ($\mu = 1.0, \beta = 0.3$) conforme a medida de acessibilidade para redes rodoviárias do (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Figura extraída de (ARRUDA *et al.*, 2014).

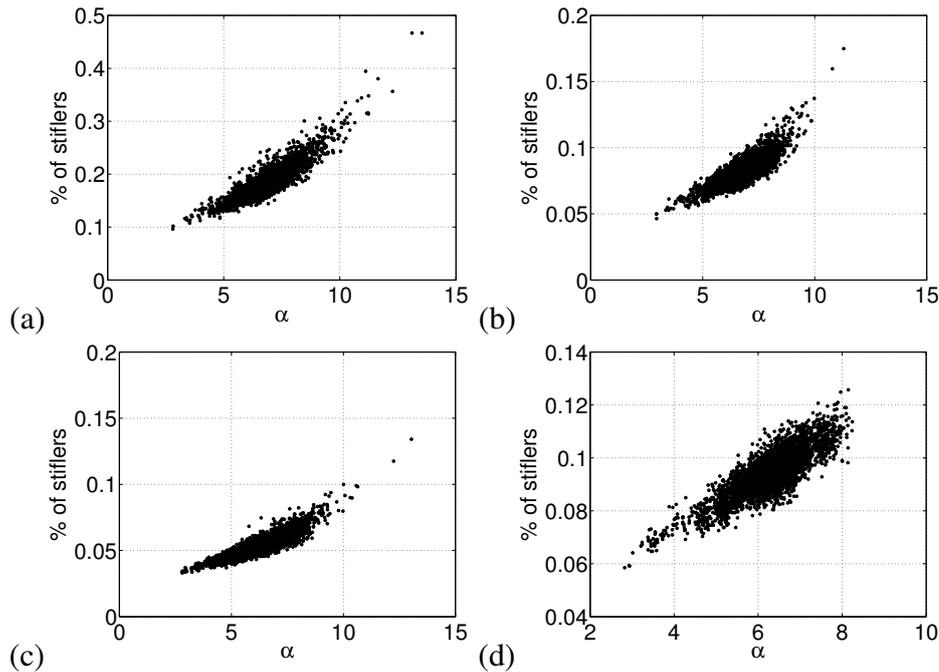


Figura 22 – A porcentagem de *stiflers* no modelo de rumor MT ($\delta = 1.0, \lambda = 0.3$) conforme medida de acessibilidade para redes rodoviárias (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Figura extraída de (ARRUDA *et al.*, 2014).

5.6 Rede não espacial

A Tabela 3 mostra as médias dos valores das medidas calculadas para as redes BA e as redes sociais. Assim como a Tabela 5 mostra o coeficiente de correlação proposto por Spearman calculado entre as métricas de centralidade e o percentual final de *stiflers* ou recuperados nas dinâmicas rumor e epidemia, respectivamente. Para o modelo SIR, as métricas *k-shell* e coenctividade são as mais relacionadas com o percentual final de recuperados. Assim, os principais *hubs* nas redes sociais estão localizadas no centro da rede, porque eles têm os maiores valores de *k-core* (ou maior *coreness*), sugerindo que tais redes não tendem a apresentar centros periféricos.

Esses resultados corroboram com a análise de propagação apresentada por Kitsak *et al.* (2010) e com estudo da difusão de rumores em (BORGE-HOLTHOEFER; MORENO, 2012). Para a medida de acessibilidade, a maior relação Spearman aparece nas redes BA e de blogs de políticos (com $\beta = 0,3$), porém os valores da relação são muito próximos aos obtidos para a conectividade e *k-shell*. Todas as demais medidas tiveram menores coeficientes de relação (Spearman) do que as métricas *k-shell*, α e *k*.

Com relação à dinâmica de rumores, os casos das simulações usando CP e TP apresentam resultados diferentes. No primeiro caso, as medidas de autovetor e acessibilidade estão fortemente correlacionadas com a fração final da *stiflers*, enquanto que, para o segundo caso, a medida *closeness centrality* e o grau médio dos vizinhos mostraram os maiores coeficientes

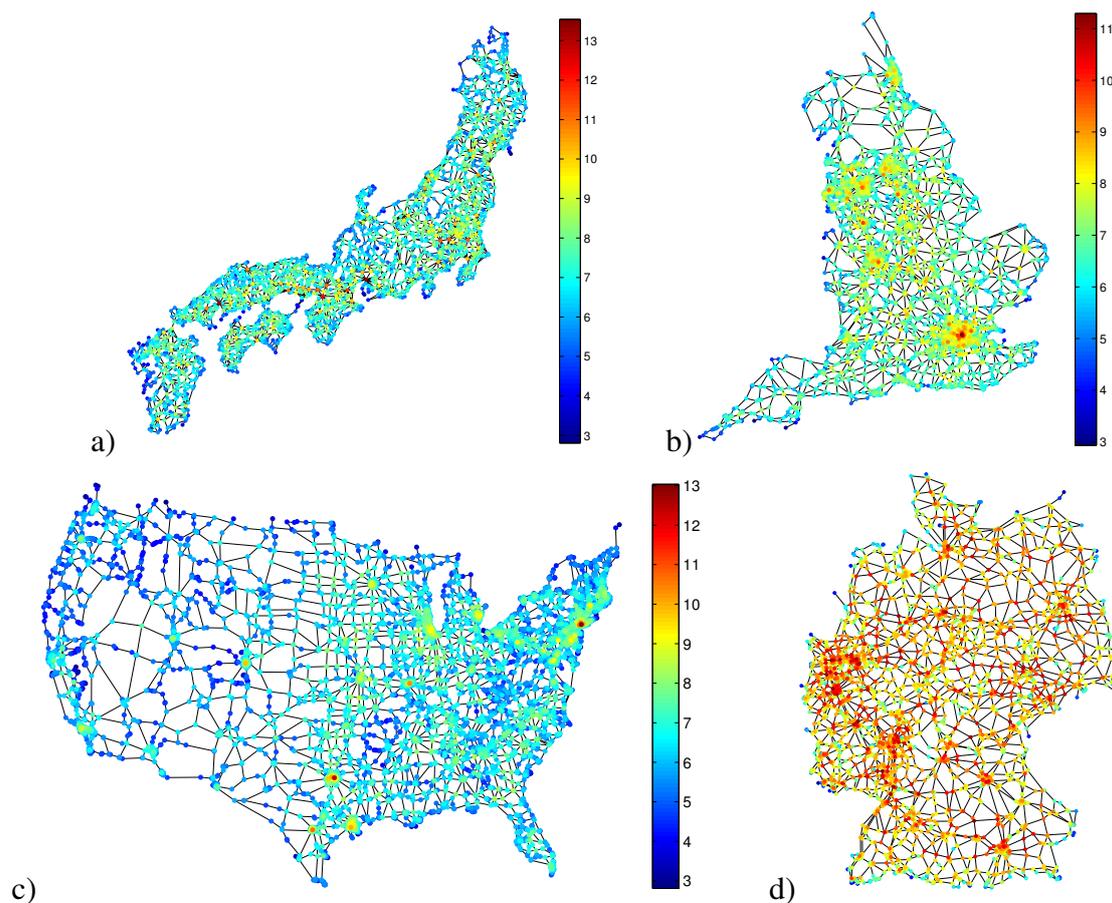


Figura 23 – Grafo das redes de rodovias do (a) Japão, (b) Inglaterra, (c) Estados Unidos da América e (d) Alemanha. Pontos em vermelho tem maior valor de acessibilidade. Figura extraída de (ARRUDA *et al.*, 2014).

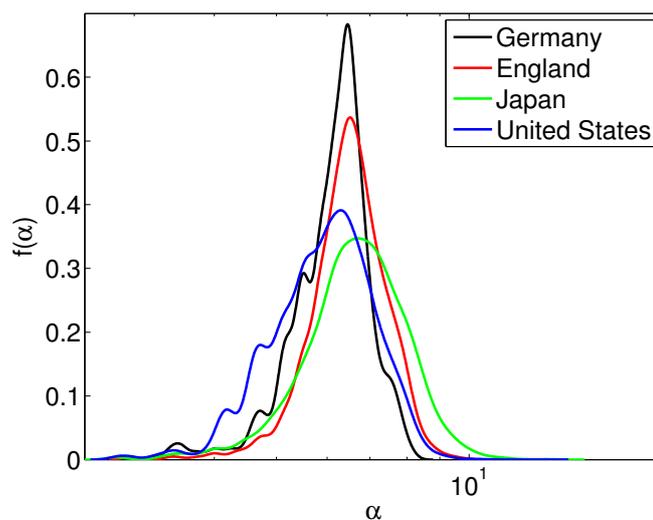


Figura 24 – Distribuição de probabilidade da medida de acessibilidade nas redes de rodovias de Japão, Alemanha, Estados Unidos e Inglaterra. Figura extraída de (ARRUDA *et al.*, 2014).

de correlação. Observando o caso TP de taxa $\lambda = 1$, e quando a taxa de propagação for alta, a conectividade média dos vizinhos estará mais relacionado com o percentual final de *contidos*. Porém, considerando taxas de propagação inferiores, a medida de distância de um vértice para o restante da rede é a medida mais crítica. Este evento fica mais fácil de observar na Tabela 5. Se r apresenta relações mais altas em taxas de difusão maiores, enquanto que a medida de proximidade é mais relacionada quando as taxas de difusão são menores. Essa avaliação mostra que os menores caminhos são mais importantes para a difusão de informação. Além disso, as medidas *k-core* e grau não foram encontrados para exibir uma forte correlação com a fração final *stiflers*, reforçando os resultados em (BORGE-HOLTHOEFER; MORENO, 2012). Por fim, notamos que mesmo em desacordo com casos observados anteriormente, para a difusão de rumores em redes não espaciais, não existe uma única medida que produzisse os maiores coeficientes de relação para todas os grafos analisados. Especificamente, a medida de acessibilidade não parece ser tão diferenciada, como anteriormente, muito provável porque, como vimos na Figura 25, as distribuições da medida acessibilidade em grafos não espaciais não são simétricos, com valores médios diferentes e uma distribuição de cauda estendida.

5.7 Conclusões

Neste capítulo analisamos a correlação entre a centralidade de um vértice e o resultado da dinâmica dos processos de epidemia e de rumor iniciados neste vértice por meio de inúmeras simulações numéricas em grafos. Nós consideramos oito métricas de centralidade da rede espaciais e não-espaciais. Propusemos uma generalização da medida de acessibilidade introduzida em (TRAVENÇOLO; F. Costa, 2008), que permite a quantificação do potencial de cada nó no acesso a outros nós de uma forma equilibrada e homogênea. Tal generalização leva em conta as caminhadas para todos os valores de comprimento h ponderando-se pelo inverso do fatorial

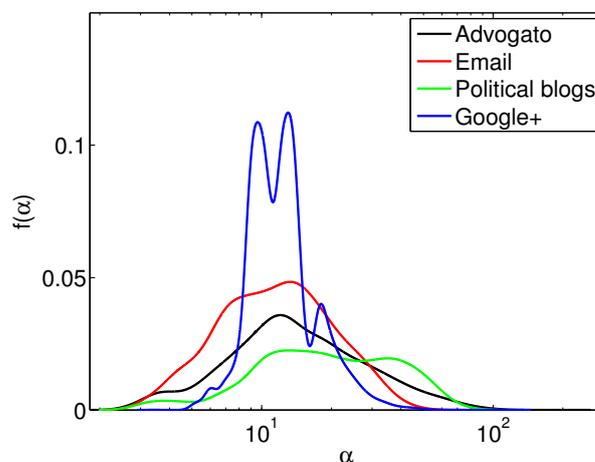


Figura 25 – Distribuição de probabilidade da medida de acessibilidade nas redes sociais Advogato, email, blogs políticos e Google+. Figura extraída de (ARRUDA *et al.*, 2014).

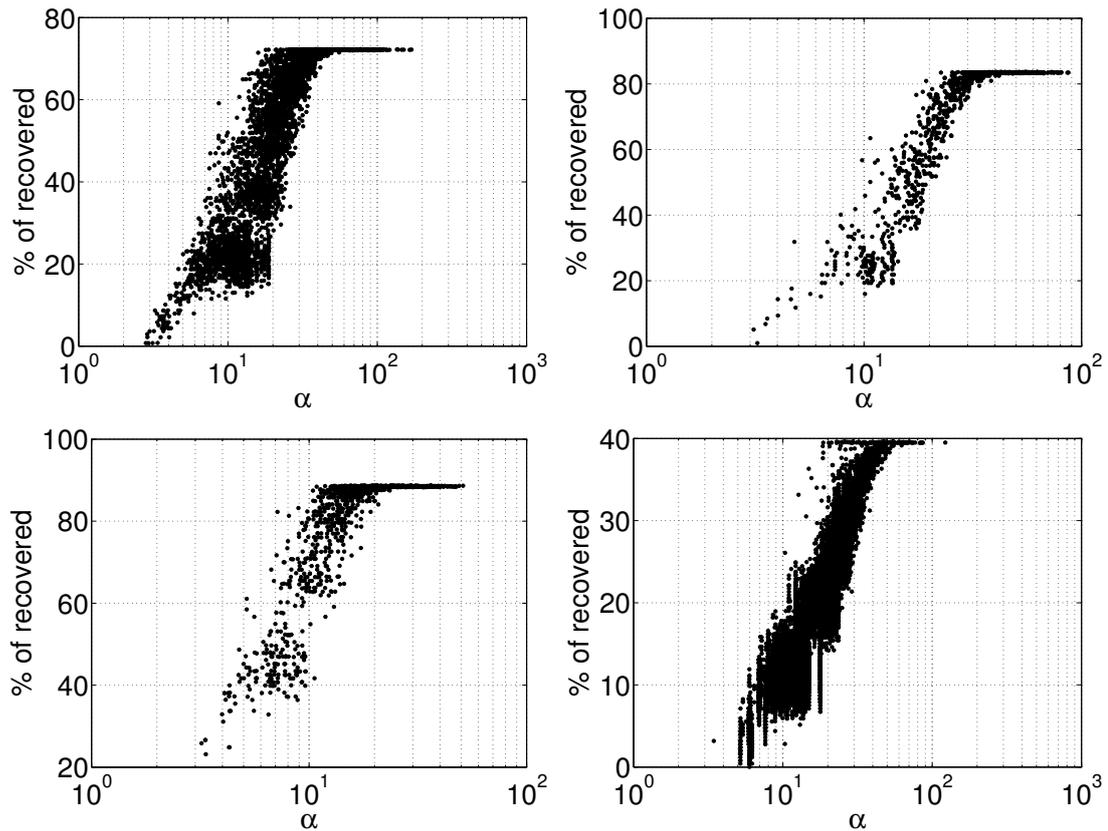


Figura 26 – Variação da medida de acessibilidade e o percentual de agentes recuperados no modelo de propagação SIR ($\beta = 0.3, \mu = 1.0$) para os grafos (a) Advogado, (b) de políticos, (c) mensagens, (d) +Google. Figura extraída de (ARRUDA *et al.*, 2014).

destes cumprimentos.

Os nossos resultados mostraram que a acessibilidade generalizada é a melhor métrica para medir a capacidade de propagação de um nó em redes espaciais. Em redes não espaciais, as melhores relações entre as medidas de centralidade e o resultado da dinâmica depende qual processo foi utilizado. Desta forma, a conectividade e a medida *k-shell* são as mais indicadas quando se trata de análise de epidemias, confirmando a análise de (KITSACK *et al.*, 2010). Porém, estas mesmas medidas não são as mais adequadas quando se utiliza um modelo de rumor. De fato, para este último caso, o grau médio dos vizinhos, a medida *closeness centrality* e a acessibilidade dão maiores correlações.

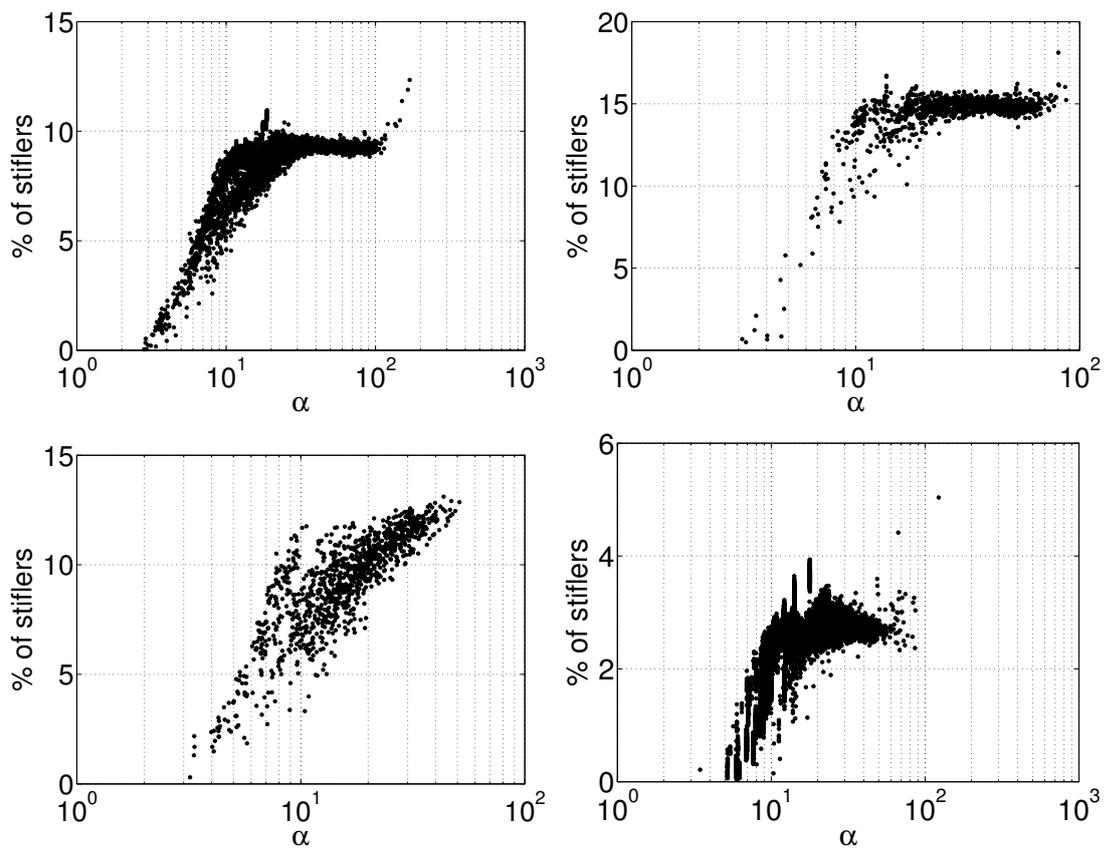


Figura 27 – Variação da medida de acessibilidade e o percentual de stiflers no modelo de rumores MT ($\delta = 1.0, \lambda = 0.3$) para os grafos (a) Advogado, (b) de políticos, (c) mensagens, (d) +Google. Figura extraída de (ARRUDA *et al.*, 2014).

Tabela 4 – Relação entre centralidade e o percentual de agentes recuperados (SIR) ou o percentual de *stiflers* (MT). As medidas são conectividade (k), coeficiente de clustering (cc), *betweenness* (B), média do grau dos vizinhos (r), Page Rank (π), autovetores (x), k-shell (k_c), proximidade (C) e acessibilidade (α). Em negrito, os maiores valores.

Process	Rates	Network	k	cc	B	r	π	x	k_c	C	α
SIR	$\beta = 0.8, \mu = 1.0$	Japan	0.40	0.11	0.24	0.26	0.27	0.27	0.35	0.47	0.47
		England	0.55	0.10	0.26	0.38	0.30	-0.04	0.21	0.27	0.58
		US	0.60	0.25	0.19	0.53	0.28	0.49	0.26	0.41	0.73
		Germany	0.54	0.05	0.42	0.35	0.20	0.22	0.19	0.34	0.63
		SpatialSF	0.81	0.25	0.65	-0.04	0.75	0.35	-	0.32	0.60
		Waxman	0.74	0.34	0.47	0.29	0.62	0.05	0.70	0.19	0.73
	$\beta = 0.3, \mu = 1.0$	Japan	0.65	0.30	0.31	0.65	0.37	0.36	0.65	0.41	0.79
		England	0.68	0.13	0.27	0.61	0.31	-0.07	0.38	0.21	0.76
		US	0.77	0.38	0.10	0.68	0.38	0.59	0.37	0.14	0.86
		Germany	0.69	0.08	0.46	0.42	0.31	0.22	0.19	0.25	0.74
		SpatialSF	0.70	0.39	0.70	0.46	0.49	0.72	-	0.66	0.91
		Waxman	0.68	0.30	0.42	0.58	0.45	0.04	0.72	0.30	0.81
MT TP	$\lambda = 0.8, \delta = 1.0$	Japan	0.50	-0.08	0.43	0.48	0.29	0.33	0.45	0.42	0.81
		England	0.54	-0.20	0.43	0.49	0.22	-0.03	0.34	0.32	0.85
		US	0.67	0.11	0.22	0.62	0.30	0.55	0.33	0.24	0.90
		Germany	0.57	-0.24	0.62	0.41	0.20	0.24	0.18	0.33	0.88
		SpatialSF	0.63	0.29	0.71	0.55	0.40	0.75	-	0.70	0.94
		Waxman	0.56	0.06	0.52	0.46	0.37	0.04	0.58	0.45	0.76
	$\lambda = 0.8, \delta = 0.3$	Japan	0.17	-0.02	0.22	0.23	0.04	0.26	0.21	0.66	0.35
		England	0.32	-0.07	0.31	0.38	0.05	0.05	0.26	0.60	0.53
		US	0.26	0.00	0.25	0.29	0.07	0.08	0.12	0.83	0.43
		Germany	0.29	-0.12	0.46	0.28	0.01	0.45	0.17	0.67	0.52
		SpatialSF	0.40	0.16	0.43	0.28	0.27	0.41	-	0.37	0.53
		Waxman	0.61	0.13	0.51	0.38	0.47	0.06	0.62	0.31	0.74
	$\lambda = 0.3, \delta = 1.0$	Japan	0.77	0.22	0.43	0.61	0.54	0.25	0.59	0.28	0.88
		England	0.77	0.03	0.34	0.53	0.47	-0.07	0.30	0.16	0.83
		US	0.84	0.32	0.19	0.63	0.50	0.56	0.35	0.13	0.91
		Germany	0.73	0.01	0.45	0.40	0.39	0.17	0.19	0.20	0.79
		SpatialSF	0.34	0.32	0.53	0.71	0.12	0.89	-	0.84	0.77
		Waxman	0.84	0.30	0.56	0.59	0.64	0.06	0.77	0.25	0.94
	$\lambda = 0.3, \delta = 0.3$	Japan	0.37	0.00	0.32	0.50	0.13	0.35	0.42	0.49	0.68
		England	0.42	-0.09	0.34	0.52	0.07	0.01	0.37	0.38	0.71
US		0.54	0.12	0.15	0.64	0.16	0.54	0.33	0.28	0.80	
Germany		0.42	-0.20	0.54	0.41	0.06	0.29	0.18	0.39	0.73	
SpatialSF		0.42	0.31	0.53	0.62	0.19	0.71	-	0.65	0.84	
Waxman		0.44	0.08	0.41	0.46	0.25	0.09	0.51	0.55	0.64	
MT CP	$\lambda = 0.8, \delta = 1.0$	Japan	0.42	0.05	0.30	0.57	0.16	0.32	0.48	0.42	0.73
		England	0.43	-0.10	0.32	0.56	0.08	-0.05	0.36	0.32	0.73
		US	0.57	0.16	0.13	0.682	0.17	0.55	0.34	0.24	0.82
		Germany	0.45	-0.18	0.52	0.44	0.07	0.26	0.18	0.35	0.75
		SpatialSF	0.26	0.27	0.42	0.69	0.02	0.75	-	0.71	0.73
		Waxman	0.50	0.13	0.39	0.57	0.27	0.04	0.59	0.38	0.72
	$\lambda = 0.8, \delta = 0.3$	Japan	0.17	0.01	0.18	0.28	0.01	0.27	0.26	0.67	0.36
		England	0.24	0.00	0.24	0.37	-0.03	0.16	0.29	0.70	0.43
		US	0.29	0.04	0.20	0.36	0.04	0.20	0.16	0.74	0.47
		Germany	0.20	-0.07	0.41	0.24	-0.05	0.51	0.16	0.81	0.40
		SpatialSF	0.28	0.22	0.36	0.46	0.10	0.48	-	0.44	0.61
		Waxman	0.54	0.16	0.44	0.55	0.34	0.10	0.64	0.44	0.76
	$\lambda = 0.3, \delta = 1.0$	Japan	0.60	0.18	0.36	0.66	0.34	0.28	0.56	0.31	0.82
		England	0.58	-0.02	0.31	0.60	0.25	-0.10	0.31	0.18	0.77
		US	0.68	0.26	0.12	0.71	0.30	0.56	0.35	0.14	0.85
		Germany	0.55	-0.10	0.45	0.47	0.20	0.17	0.18	0.21	0.74
		SpatialSF	0.23	0.27	0.42	0.71	0.01	0.81	-	0.77	0.70
		Waxman	0.68	0.24	0.45	0.67	0.45	0.06	0.72	0.26	0.87
	$\lambda = 0.3, \delta = 0.3$	Japan	0.37	0.03	0.27	0.54	0.12	0.33	0.45	0.46	0.68
		England	0.40	-0.08	0.32	0.53	0.06	-0.01	0.36	0.36	0.69
US		0.54	0.14	0.13	0.65	0.14	0.53	0.33	0.27	0.79	
Germany		0.40	-0.20	0.52	0.43	0.03	0.27	0.18	0.39	0.72	
SpatialSF		0.27	0.27	0.43	0.67	0.03	0.71	-	0.66	0.75	
Waxman		0.44	0.10	0.38	0.51	0.23	0.05	0.53	0.46	0.65	

Tabela 5 – Relaçã entre centralidade e o percentual de stiflers (MT) ou o percentual de recuperados (SIR). As medidas são conectividade (k), coeficiente de clustering (cc), *betweenness* (B), média do grau dos vizinhos (r), Page Rank (π), autovetores (x), k-shell (k_c), proximidade (C) e acessibilidade (α). Em negrito, os maiores valores.

Processo	Taxa	Rede	k	cc	B	r	π	x	k_c	C	α	
SIR	$\mu = 1.0, \beta = 0.8$	Advogato	0.76	0.47	0.66	0.14	0.71	0.65	0.76	0.63	0.70	
		email	0.67	0.41	0.57	0.09	0.64	0.58	0.67	0.57	0.63	
		polblogs	0.57	0.29	0.51	0.04	0.55	0.55	0.57	0.55	0.57	
		Google+ BA	0.81	0.67	0.61	0.17	0.64	0.50	0.81	0.48	0.63	
				0.19	0.13	0.39	0.48	-0.02	0.45	-	0.49	0.60
	$\beta = 0.3, \mu = 1.0$	advogato	0.97	0.40	0.85	0.19	0.92	0.88	0.97	0.84	0.92	
		email	0.97	0.34	0.85	0.25	0.94	0.88	0.96	0.89	0.94	
		polblogs	0.89	0.25	0.78	-0.08	0.86	0.85	0.89	0.82	0.88	
Google+ BA		0.81	0.67	0.61	0.23	0.60	0.56	0.81	0.54	0.68		
			0.18	0.18	0.48	0.73	-0.10	0.66	-	0.72	0.77	
MT-TP	$\lambda = 0.8, \delta = 1.0$	advogato	0.23	0.17	0.26	0.55	0.17	0.38	0.22	0.43	0.35	
		email	0.62	0.19	0.61	0.13	0.60	0.56	0.58	0.60	0.64	
		polblogs	-0.16	-0.09	-0.06	0.35	-0.17	-0.16	-0.21	-0.07	-0.13	
		Google+ BA	0.04	0.03	0.00	0.78	-0.07	0.27	0.04	0.25	0.40	
				0.16	0.19	0.48	0.76	-0.12	0.70	-	0.75	0.78
	$\lambda = 0.8, \delta = 0.3$	Advogato	0.05	0.15	0.10	0.51	0.02	0.20	0.05	0.27	0.16	
		emails	0.29	0.22	0.30	0.06	0.28	0.24	0.26	0.28	0.30	
		políticos	-0.37	-0.01	-0.26	0.48	-0.37	-0.34	-0.39	-0.22	-0.33	
		+Google B A	0.00	0.03	-0.05	0.69	0.01	0.11	0.004	0.09	0.27	
				0.14	0.19	0.47	0.77	-0.11	0.79	-	0.82	0.72
	$\delta = 1.0, \lambda = 0.3$	Advogato	0.54	0.11	0.47	0.64	0.45	0.74	0.55	0.76	0.73	
		email	0.77	0.03	0.71	0.59	0.70	0.89	0.76	0.91	0.89	
		polblogs	0.19	0.05	0.19	0.41	0.17	0.19	0.16	0.31	0.26	
		Google+ BA	0.20	0.17	0.14	0.84	-0.12	0.63	0.20	0.61	0.65	
				0.35	0.11	0.46	0.34	0.20	0.47	-	0.48	0.51
	$\lambda = 0.3, \delta = 0.3$	advogato	0.36	0.19	0.31	0.57	0.29	0.53	0.37	0.57	0.52	
email		0.70	0.19	0.62	0.35	0.65	0.71	0.70	0.74	0.77		
polblogs		-0.16	0.09	-0.12	0.52	-0.18	-0.13	-0.18	-0.02	-0.10		
Google+ BA		0.14	0.16	0.05	0.76	-0.01	0.36	0.15	0.34	0.47		
			0.33	0.19	0.59	0.67	0.06	0.67	-	0.73	0.85	
MT CP	$\lambda = 0.8, \delta = 1.0$	advogato	0.47	0.14	0.35	0.52	0.36	0.63	0.50	0.62	0.65	
		email	0.69	0.19	0.56	0.57	0.61	0.81	0.73	0.79	0.81	
		polblogs	0.29	0.14	0.21	0.26	0.25	0.28	0.29	0.34	0.34	
		Google+ BA	0.40	0.32	0.31	0.55	-0.12	0.84	0.40	0.76	0.75	
				0.56	0.19	0.78	0.63	0.30	0.74	-	0.80	0.94
	$\delta = 0.3, \lambda = 0.8$	Advogato	0.29	0.14	0.21	0.35	0.21	0.38	0.31	0.39	0.41	
		email	0.45	0.22	0.36	0.32	0.40	0.48	0.47	0.50	0.52	
		polblogs	0.01	0.14	-0.01	0.26	-0.03	0.02	0.00	0.07	0.05	
		Google+ BA	0.32	0.27	0.23	0.52	-0.10	0.64	0.33	0.60	0.64	
				0.24	0.21	0.58	0.80	-0.02	0.87	-	0.91	0.79
	$\lambda = 0.3, \delta = 1.0$	advogato	0.52	0.08	0.39	0.59	0.40	0.72	0.55	0.71	0.73	
		email	0.64	0.05	0.55	0.68	0.55	0.83	0.67	0.82	0.80	
		polblogs	0.51	0.14	0.39	0.25	0.46	0.52	0.51	0.56	0.57	
		Google+ BA	0.37	0.28	0.32	0.55	-0.15	0.89	0.38	0.84	0.75	
				0.82	0.10	0.67	-0.01	0.74	0.31	-	0.34	0.52
	$\lambda = 0.3, \delta = 0.3$	advogato	0.46	0.15	0.34	0.50	0.35	0.60	0.48	0.60	0.63	
email		0.68	0.23	0.55	0.50	0.60	0.76	0.71	0.76	0.79		
polblogs		0.19	0.14	0.13	0.27	0.15	0.18	0.18	0.24	0.23		
Google+ BA		0.39	0.32	0.30	0.55	-0.11	0.80	0.40	0.74	0.74		
			0.67	0.19	0.79	0.49	0.43	0.66	-	0.72	0.89	

CONCLUSÃO

Neste trabalho abordamos dois métodos principais na análise de vértices influentes em redes complexas ao simularmos dinâmicas de propagação como epidemias ou rumores. Para a primeira abordagem, Capítulo 4, baseada no algoritmo de maximização da esperança e um conjunto de medidas de centralidade, verificamos que ela evita a arbitrariedade na escolha de uma única medida de centralidade, uma vez que não existe uma métrica que seja suficientemente geral para quantificar o papel central de um nó. Os vértices com os maiores valores de medidas de centralidade individuais mostraram ser os mais influentes propagadores de doenças. Verificou-se que para todas as redes, que vacinar os vértices no cluster mais central é a melhor maneira para diminuir a epidemia. Os resultados mostram que o cluster que contém os nós mais centrais são os propagadores mais influentes do modelo SIR. Além disso, este método proposto permite a determinação do número de propagadores influentes automaticamente.

Na segunda abordagem, Capítulo 5, procuramos identificar qual medida estrutural melhor descreve os vértices mais influentes na propagação conforme a rede utilizada. Os resultados mostraram que a acessibilidade generalizada é a melhor métrica para medir a capacidade de propagação de um nó em redes espaciais. Porém, em redes não espaciais, as maiores correlações entre uma métrica de centralidade e os valores obtidos na dinâmica vai depender do processo escolhido. Esta nova medida de acessibilidade generalizada mostrou ser a melhor forma de identificar propagadores mais importantes em grafos espaciais, logo que, alegações feitas anteriormente sobre se uma classe de vértices pode ser mais influente depende tanto da medida usada quanto do tipo de grafo em estudo.

Podemos dizer, que a identificação dos mais influentes propagadores dentro de uma rede é muito importante mas também difícil de se generalizar uma única medida para todos os tipos de redes. Esta identificação se torna mais importante, quando deixamos a etapa de simulação e passamos a nos referir a epidemias de doenças se alastrando ao redor do globo terrestre, onde a rápida identificação dos maiores propagadores serve para controlar e extinguir tais surtos. Por

isso, novos métodos e medidas são de suma importância.

Há diversas maneiras de estender a nossa análise apresentada nesse trabalho. Um estudo natural é considerar outras medidas de centralidade, bem como outros modelos de propagação de epidemias e rumores. Particularmente, iniciamos um estudo para identificar propagadores mais influentes em um modelo epidêmico com mutação (GIRVAN *et al.*, 2002; SHAO *et al.*, 2006). Esse modelo é mais realista do que os modelos tradicionais, pois o agente infeccioso pode mudar o seu código genético ao longo do tempo e infectar indivíduos que antes estavam imunizados. Na verdade, esse modelo é mais adequado para modelar doenças como o novo coronavírus, já que o vírus sofreu várias mutações e infectou indivíduos anteriormente imunizados.

O modelo que consideramos é apresentado em (SHAO *et al.*, 2006). Basicamente, a cada vértice tem associado um vetor binário de comprimento ℓ que define o seu nível de imunidade. O agente infeccioso também é representado por um vetor binário de mesmo comprimento. A infecção ocorre através da comparação entre os vetores dos indivíduos e do vírus, sendo calculada a distância entre eles usando a distância de Hamming. Particularmente, um indivíduo é suscetível se a distância entre o seu vetor e o do vírus é menor do que um certo limiar definido, $h_{min} > h_{thr}$, onde h_{thr} é o limiar. Um indivíduo infectado transmite a doença após um tempo de infecção τ_I . Da mesma forma, um agente volta ao estado suscetível após um período τ_R . Após esse tempo, os vetores binários dos vértices voltam ao estado inicial, ou seja, apresentam entradas nulas. O vírus sofre uma mutação em um de seus bits de forma aleatória, de acordo com uma probabilidade μ . Seleccionamos $\mu = 0,01$ como no artigo original (SHAO *et al.*, 2006). Como condição inicial, assumimos que todos os indivíduos começam com vetores com entradas nulas. De acordo com as mutações, os vírus mudam os seus vetores e começam a infectar os indivíduos. Cada indivíduo infectado, perde a memória imunológica ao longo do tempo. A figura 28 mostra a evolução temporal do modelo para diferentes conjuntos de parâmetros. Conforme observamos, esse modelo consegue capturar tanto as dinâmicas SIR como SIS.

Analisando esse modelo em termos da identificação de propagadores influentes, observamos que não é possível tal identificação usando medidas de centralidade. No entanto, mais estudos devem ser feitos, já que esse modelo é mais realista do que os modelos tradicionais. Sugerimos esse estudo como um trabalho futuro.

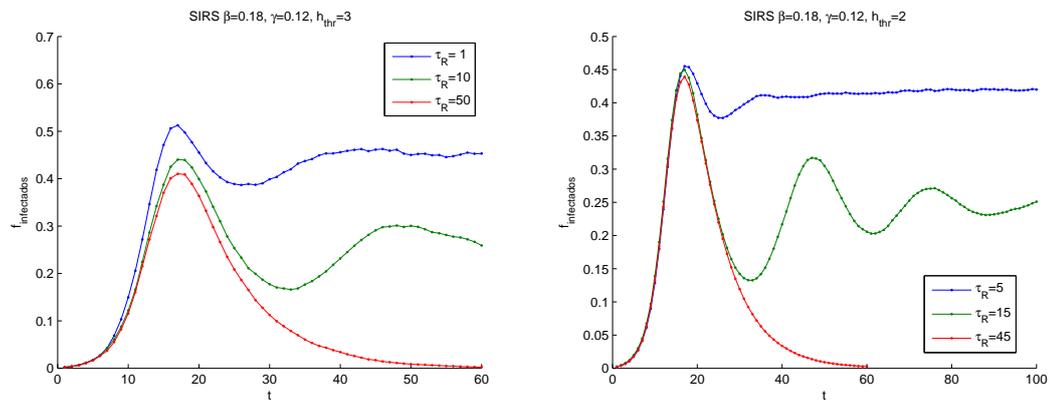


Figura 28 – Fração de infectados ao longo do tempo usando diferentes valores das medidas τ_R e h_{thr} em redes mundo pequeno.

Fonte: Elaborada pelo autor.

REFERÊNCIAS

ADAMIC, L. A.; GLANCE, N. The political blogosphere and the 2004 US election: divided they blog. In: ACM. **Proceedings of the 3rd international workshop on Link discovery**. [S.l.], 2005. p. 36–43. Citado nas páginas 51 e 68.

ADVOGATO network dataset – KONECT. 2014. Disponível em: <<http://konect.uni-koblenz.de/networks/advogato>>. Citado na página 68.

ALBERT, R.; JEONG, H.; BARABÁSI, A.-L. Diameter of the World Wide Web. **Nature**, v. 401, p. 130–131, 1999. Citado na página 39.

ARRUDA, G. F. de; BARBIERI, A. L.; RODRÍGUEZ, P. M.; RODRIGUES, F. A.; MORENO, Y.; COSTA, L. d. F. Role of centrality for the identification of influential spreaders in complex networks. **Phys. Rev. E**, American Physical Society, v. 90, p. 032812, Sep 2014. Disponível em: <[doi:10.1103/PhysRevE.90.032812](https://doi.org/10.1103/PhysRevE.90.032812)>. Citado nas páginas 16, 17, 60, 61, 63, 69, 70, 72, 73, 74, 75, 76, 77 e 78.

BABAEI, A.; AHMADI, M.; JAFARI, H.; LIYA, A. A mathematical model to examine the effect of quarantine on the spread of coronavirus. **Chaos, Solitons & Fractals**, Elsevier, v. 142, p. 110418, 2021. Citado na página 25.

BARABÁSI, A.-L. **Linked: how everything is connected to everything else and what it means for business, science, and everyday life**. New York: Penguin Group, 2003. Citado na página 25.

BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509–12, 1999. Citado nas páginas 34, 39 e 40.

BARABÁSI, A.-L.; JEONG, H.; RAVASZ, R.; NÉDA, Z.; VICSEK, T.; SCHUBERT, A. On the topology of the scientific collaboration networks. **Physica A: statistical mechanics and its applications**, v. 311, p. 590–614, 2002. Citado na página 39.

BARRAT, A.; BARTHLEMY, M.; VESPIGNANI, A. **Dynamical processes on complex networks**. New York, NY, USA: Cambridge University Press, 2008. Citado nas páginas 35 e 46.

BARTHÉLEMY, M. Crossover from scale-free to spatial networks. **EPL (Europhysics Letters)**, v. 63, n. 6, p. 915, 2003. Citado na página 41.

BHATIA, R. **Matrix analysis**. New York: Springer Verlag, 1997. v. 169. Citado na página 60.

BISHOP, C. M. *et al.* **Pattern recognition and machine learning**. [S.l.]: springer New York, 2006. v. 4. Citado na página 51.

BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D. Complex networks: Structure and dynamics. **Physics Reports**, Elsevier, v. 424, n. 4, p. 175–308, 2006. Citado nas páginas 33, 35, 37, 55 e 59.

BORGE-HOLTHOEFER, J.; MORENO, Y. Absence of influential spreaders in rumor dynamics. **Physical Review E**, APS, v. 85, n. 2, p. 026116, 2012. Citado nas páginas 49, 59, 60, 71, 72, 74 e 76.

BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. **Computer networks and ISDN systems**, Elsevier, v. 30, n. 1-7, p. 107–117, 1998. Citado na página 35.

CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. **Reviews of Modern Physics**, APS, v. 81, n. 2, p. 591, 2009. Citado nas páginas 46 e 47.

CHEN, Q.; CHANG, H.; GOVINDAN, R.; JAMIN, S. The origin of power laws in Internet topologies revisited. **INFOCOM 2002 Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies**, v. 2, p. 608–617, 2002. Disponível em: <doi:10.1109/INFCOM.2002.1019306>. Citado na página 34.

COSTA, L. d. F.; Cesar Jr, R. **Shape analysis and classification: theory and practice**. Florida: CRC Press, Inc., 2000. Citado na página 68.

COSTA, L. d. F.; RODRIGUES, F. A.; HILGETAG, C. C.; KAISER, M. Beyond the average: Detecting global singular nodes from local features in complex networks. **Europhysics Letters**, IOP Publishing, v. 87, n. 1, p. 18008, 2009. Citado na página 54.

COSTA, L. F.; RODRIGUES, F. A. Seeking for simplicity in complex networks. **Europhysics Letters**, IOP Publishing, v. 85, n. 4, p. 48001, 2009. Citado na página 54.

DALEY, D. J.; GANI, J.; GANI, J. M. **Epidemic modelling: an introduction**. Cambridge: Cambridge University Press, 2001. v. 15. Citado na página 46.

DAVIS, G.; YOO, M.; BAKER, W. The small world of the American corporate elite, 1982-2001. **Strategic Organization**, Sage Publications, v. 1, n. 3, p. 301, 2003. ISSN 1476-1270. Citado na página 34.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society**, JSTOR, p. 1–38, 1977. Citado na página 51.

DIETZ, K.; HEESTERBEEK, J. Daniel bernoulli's epidemiological model revisited. **Mathematical biosciences**, Elsevier, v. 180, n. 1-2, p. 1–21, 2002. Citado na página 43.

DOUGHERTY, E. R. Book. **An introduction to morphological image processing**. USA: SPIE Optical Engineering Press Bellingham, Wash., 1992. xii, 161 p. : p. ISBN 081940845. Citado na página 68.

ERKOL, Ş.; CASTELLANO, C.; RADICCHI, F. Systematic comparison between methods for the detection of influential spreaders in complex networks. **Scientific reports**, Nature Publishing Group, v. 9, n. 1, p. 1–11, 2019. Citado na página 50.

ESTRADA, E.; HATANO, N. Communicability in complex networks. **Physical Review E**, APS, v. 77, n. 3, p. 036111, 2008. Citado na página 61.

ESTRADA, E.; HATANO, N.; BENZI, M. The Physics of Communicability in Complex Networks. **CoRR**, abs/1109.2950, 2011. Citado na página 61.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the Internet topology. **Computer Communication Review**, v. 29, n. 4, p. 251–262, 1999. Citado na página 39.

FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, JSTOR, v. 40, n. 1, p. 35–41, 1977. ISSN 0038-0431. Citado na página 33.

GIRVAN, M.; CALLAWAY, D. S.; NEWMAN, M. E.; STROGATZ, S. H. Simple model of epidemics with pathogen mutation. **Physical Review E**, APS, v. 65, n. 3, p. 031915, 2002. Citado na página 82.

GOOGLE+ network dataset – KONECT. 2014. Disponível em: <<http://konect.uni-koblenz.de/networks/ego-gplus>>. Citado na página 68.

GROSSMAN, J.; ION, P. On a portion of the well-known collaboration graph. **Congressus Numerantium**, CiteSeer, p. 129–132, 1995. ISSN 0384-9864. Citado na página 34.

GUIMERA, R.; DANON, L.; DÍAZ-GUILERA, A.; GIRALT, F.; ARENAS, A. Self-similar community structure in a network of human interactions. **Physical Review E**, v. 68, p. 065103, 2003. Citado nas páginas 51 e 68.

HILL, M. O. Diversity and evenness: a unifying notation and its consequences. **Ecology**, Eco Soc America, v. 54, n. 2, p. 427–432, 1973. Citado na página 61.

JEONG, H.; MASON, S.; BARABÁSI, A.-L.; OLTVAI, Z. Lethality and centrality in protein networks. **Nature**, Nature Publishing Group, v. 411, n. 6833, p. 41–42, 2001. Citado nas páginas 34 e 39.

JEONG, H.; TOMBOR, B.; ALBERT, R.; OLTVAI, Z.; BARABÁSI, A.-L. The large-scale organization of metabolic networks. **Nature**, Nature Publishing Group, v. 407, n. 6804, p. 651–654, 2000. Citado na página 39.

JOST, L. Entropy and diversity. **Oikos**, Wiley Online Library, v. 113, n. 2, p. 363–375, 2006. Citado na página 61.

KAMADA, T.; KAWAI, S. An algorithm for drawing general undirected graphs. **Inf. Process. Lett.**, Elsevier North-Holland, Inc., Amsterdam, The Netherlands, The Netherlands, v. 31, n. 1, p. 7–15, abr. 1989. ISSN 0020-0190. Disponível em: <[http://dx.doi.org/10.1016/0020-0190\(89\)90102-6](http://dx.doi.org/10.1016/0020-0190(89)90102-6)>. Citado nas páginas 15, 16, 55 e 57.

KITSAK, M.; GALLOS, L.; HAVLIN, S.; LILJEROS, F.; MUCHNIK, L.; STANLEY, H.; MAKSE, H. Identification of influential spreaders in complex networks. **Nature Physics**, Nature Publishing Group, v. 6, n. 11, p. 888–893, 2010. Citado nas páginas 25, 36, 49, 55, 57, 59, 60, 71, 72, 74 e 77.

LEVEQUE, R. J. **Finite difference methods for ordinary and partial differential equations - steady-state and time-dependent problems**. [S.l.]: SIAM, 2007. I–XV, 1–341 p. ISBN 978-0-89871-629-0. Citado na página 65.

LÜ, L.; CHEN, D.; REN, X.-L.; ZHANG, Q.-M.; ZHANG, Y.-C.; ZHOU, T. Vital nodes identification in complex networks. **Physics Reports**, Elsevier, v. 650, p. 1–63, 2016. Citado na página 25.

MARTINEZ, N. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. **Ecological Monographs**, JSTOR, v. 61, n. 4, p. 367–392, 1991. ISSN 0012-9615. Citado na página 34.

MASSA, P.; SALVETTI, M.; TOMASONI, D. Bowling Alone and Trust Decline in Social Network Sites. In: **Proc. Int. Conf. Dependable, Autonomic and Secure Computing**. [S.l.: s.n.], 2009. p. 658–663. Citado na página 68.

MAY, A.; MILNE, D. Effects of alternative road pricing systems on network performance. **Transportation Research Part A: policy and practice**, Elsevier, v. 34, n. 6, p. 407–436, 2000. ISSN 0965-8564. Citado na página 31.

MCAULEY, J.; LESKOVEC, J. Learning to Discover Social Circles in Ego Networks. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2012. p. 548–556. Citado na página 68.

MIEGHEM, P. V. **Graph Spectra for Complex Networks**. New York, NY, USA: Cambridge University Press, 2011. ISBN 9780521194587. Citado na página 66.

NEWMAN, M. Power laws, Pareto distributions and Zipf’s law. **Contemporary physics**, Taylor & Francis, v. 46, n. 5, p. 323–351, 2005. Citado na página 40.

_____. **Networks: an introduction**. New York: Oxford University Press, Inc., 2010. Citado nas páginas 31, 36, 50 e 59.

NEWMAN, M. E. Assortative mixing in networks. **Physical Review Letters**, APS, v. 89, n. 20, p. 208701, 2002. ISSN 1079-7114. Citado nas páginas 34, 35 e 41.

NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 2, p. 404–409, 2001. Citado na página 34.

PARETO, V. **The mind and society: trattato di sociologia generale**. New York: Harcourt Brace and Company, 1935. Citado na página 40.

PARZEN, E. On estimation of a probability density function and mode. **The Annals of Mathematical Statistics**, JSTOR, v. 33, n. 3, p. 1065–1076, 1962. Citado na página 53.

PASTOR-SATORRAS, R.; CASTELLANO, C. Topological structure and the h index in complex networks. **Physical Review E**, APS, v. 95, n. 2, p. 022301, 2017. Citado na página 59.

PASTOR-SATORRAS, R.; CASTELLANO, C.; MIEGHEM, P. V.; VESPIGNANI, A. Epidemic processes in complex networks. **Reviews of Modern Physics**, APS, v. 87, n. 3, p. 925, 2015. Citado na página 25.

POUX-MÉDARD, G.; PASTOR-SATORRAS, R.; CASTELLANO, C. Influential spreaders for recurrent epidemics on networks. **Physical Review Research**, APS, v. 2, n. 2, p. 023332, 2020. Citado na página 59.

RADICCHI, F.; CASTELLANO, C. Leveraging percolation theory to single out influential spreaders in networks. **Physical Review E**, APS, v. 93, n. 6, p. 062314, 2016. Citado na página 25.

ROSS, S. M. **Introduction to probability models**. New York: Academic Press, 2007. ISBN 0125980620. Citado na página 37.

SEIDMAN, S. Network structure and minimum degree. **Social networks**, Elsevier, v. 5, n. 3, p. 269–287, 1983. Citado nas páginas 36 e 60.

SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N.; WANG, J.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome Research**, Cold Spring Harbor Lab, v. 13, n. 11, p. 2498, 2003. ISSN 1088-9051. Citado nas páginas 15, 16, 55 e 57.

SHAO, Z.-G.; TAN, Z.-J.; ZOU, X.-W.; JIN, Z.-Z. Epidemics with pathogen mutation on small-world networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier, v. 363, n. 2, p. 561–566, 2006. Citado na página 82.

SILVA, R. A. P. da; VIANA, M. P.; COSTA, L. da F. Predicting epidemic outbreak from individual features of the spreaders. **Journal of Statistical Mechanics: Theory and Experiment**, IOP Publishing, v. 2012, n. 07, p. P07005, 2012. Citado na página 59.

SNEPPEN, K.; SIMONSEN, L. Impact of superspreaders on dissemination and mitigation of covid-19. **medRxiv**, Cold Spring Harbor Laboratory Press, 2020. Citado na página 25.

TANENBAUM, A. S. **Computer Networks**. Upper Saddle River: Prentice Hall, 2003. ISBN 0130661023. Citado na página 32.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 2nd. ed. New York: Academic Press, 2003. Citado nas páginas 50, 51 e 52.

TRAUD, A.; MUCHA, P.; PORTER, M. Social structure of facebook networks. **Physica A: Statistical Mechanics and its Applications**, Elsevier, 2011. Citado na página 51.

TRAVENÇOLO, B.; VIANA, M.; COSTA, L. d. F. Border detection in complex networks. **New Journal of Physics**, IOP Publishing, v. 11, p. 063019, 2009. Citado nas páginas 60 e 73.

TRAVENÇOLO, B. A. N.; F. Costa, L. da. Accessibility in complex networks. **Physics Letters A**, Elsevier, v. 373, n. 1, p. 89–95, 2008. Citado nas páginas 60 e 76.

VÁZQUEZ, A.; PASTOR-SATORRAS, R.; VESPIGNANI, A. Large-scale topological and dynamical properties of the Internet. **Physical Review E**, APS, v. 65, n. 6, p. 66130, 2002. Disponível em: <doi:10.1103/PhysRevE.65.066130>. Citado na página 34.

WATTS, D. **Small Worlds: The Dynamics of Networks Between Order and Randomness**. Princeton: Princeton University Press, 1999. Citado na página 41.

WATTS, D.; STROGATZ, S. Collective dynamics of small-world networks. **Nature**, Nature Publishing Group, v. 393, n. 6684, p. 440–442, 1998. ISSN 0028-0836. Citado nas páginas 30, 34, 38 e 39.

WAXMAN, B. Routing of multipoint connections. **Selected Areas in Communications, IEEE Journal on**, IEEE, v. 6, n. 9, p. 1617–1622, 1988. Citado na página 41.

WITTEN, I.; FRANK, E. **Data Mining: Practical machine learning tools and techniques**. [S.l.]: Morgan Kaufmann, 2005. Citado na página 51.

WOLFE, D. A.; HOLLANDER, M. Nonparametric statistical methods. **Nonparametric statistical methods**, John Wiley New York, 1973. Citado na página 71.

ZACHARY, W. An information flow model for conflict and fission in small groups¹. **Journal of anthropological research**, v. 33, n. 4, p. 452–473, 1977. Citado nas páginas 15, 16, 30 e 61.

ZHANG, J.-X.; CHEN, D.-B.; DONG, Q.; ZHAO, Z.-D. Identifying a set of influential spreaders in complex networks. **Scientific reports**, Nature Publishing Group, v. 6, p. 27823, 2016. Citado na página 59.

