

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Abraão Guimarães Flores

**Difusão Orientada por Centralidade em Redes
Complexas Dinâmicas**

Juiz de Fora

2013

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Abraão Guimarães Flores

**Difusão Orientada por Centralidade em Redes
Complexas Dinâmicas**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientadora: Ana Paula Couto da Silva

Coorientador: Alex Borges Vieira

Juiz de Fora

2013

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Flores, Abraão Guimarães.
Difusão Orientada por Centralidade em Redes Complexas Dinâmicas / Abraão Guimarães Flores. -- 2013.
70 p. : il.

Orientador: Ana Paula Couto da Silva
Coorientador: Alex Borges Vieira
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2013.

1. Difusão. 2. Dinâmica. 3. Redes Complexas. 4. Tempo de Cobertura. I. Silva, Ana Paula Couto da, orient. II. Vieira, Alex Borges, coorient. III. Título.

Abraão Guimarães Flores

Difusão Orientada por Centralidade em Redes Complexas Dinâmicas

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 26 de Agosto de 2013.

BANCA EXAMINADORA

Prof^ª. D.Sc. Ana Paula Couto da Silva - Orientadora
Universidade Federal de Minas Gerais

Prof. D.Sc. Alex Borges Vieira
Universidade Federal de Juiz de Fora

Prof. D.Sc. Artur Ziviani
Laboratório Nacional de Computação Científica

Prof. D.Sc. Raul Fonseca Neto
Universidade Federal de Juiz de Fora

*Dedico este trabalho a minha
família pelo apoio. A minha
noiva que sempre esteve do meu
lado. A todos os professores que
me acompanharam nessa
trajetória.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, não sei o que seria de mim sem a fé que eu tenho nele. Aos meus pais, Walter e Nina, ao meu irmão Jonatan e a toda minha família que, com muito carinho e apoio, me incentivaram e não mediram esforços para que eu chegasse até esta etapa de minha vida. Agradeço à minha noiva Claudiana pelo carinho, amor e companherismo. À professora Ana Paula pela orientação e incentivo que tornaram possível a conclusão deste trabalho. Ao professor Alex pela ajuda e sabedoria compartilhadas durante esta caminhada. Agradeço também a todos os professores que participaram da minha vida acadêmica. Hoje sou o somatório do conhecimento compartilhado por cada um deles. Obrigado a todos.

*"Procure ser uma pessoa de
valor, em vez de procurar ser
uma pessoa de sucesso. O
sucesso é consequência."*

Albert Einstein

RESUMO

A dinamicidade é uma característica presente em diversos sistemas reais, tais como redes de comunicação, sociais, biológicas e tecnológicas. Processos de difusão em redes complexas podem surgir, por exemplo, em busca de dados, roteamento de dados e propagação de doenças. Desta forma, a compreensão do tempo necessário para difusão é um tema de estudo importante em redes complexas dinâmicas. Nesta dissertação é realizado um estudo de como medidas de centralidade podem ajudar na diminuição do tempo de difusão de informação em redes complexas dinâmicas. Usando dados de sistemas reais e sintéticos é mostrado que, se a dinamicidade é desconsiderada, o tempo necessário para difundir uma informação na rede é subestimado. Foram propostos algoritmos de difusão que consideram métricas de centralidade em grafos. Estes algoritmos aceleram o processo de difusão, quando comparados com algoritmos de difusão mais simples, como o *Random Walk*. Por fim, foi analisado o impacto de um modelo simples de predição de arestas nos algoritmos de difusão baseados em centralidade que foram propostos nesta dissertação.

Palavras-chave: Redes Complexas. Dinâmica. Difusão. Tempo de Cobertura.

ABSTRACT

The dynamics is a characteristic present in many real systems, such as communication networks, social, biological and technological. Diffusion processes in complex networks may arise, for example, search data, routing data and the spread of diseases. Thus, understanding the time required for diffusion is an important topic of study in dynamic complex networks. This dissertation is a study of how centrality measures can help in reducing the time information dissemination in dynamic complex networks. Using data from synthetic and real systems is shown that if the dynamics is disregarded, the time needed for spreading an information network is underestimated. Diffusion algorithms have been proposed that consider metrics of centrality in graphs. Finally, we analyze the impact of a simple model for predicting edge algorithms in diffusion based on centrality that have been proposed in this dissertation.

Keywords: Complex Networks. Dynamics. Diffusion. Cover Time.

LISTA DE FIGURAS

2.1	Influência do parâmetro p no modelo <i>Small World</i>	17
2.2	Diferença entre o modelo de grafos aleatórios e o modelo <i>Scale-Free</i>	18
2.3	Diferença entre os modelos SoA e CuT	21
2.4	Impacto dos modelos SoA e CuT no algoritmo <i>Epidêmico</i>	21
2.5	Impacto dos modelos SoA e CuT no algoritmo <i>Random Walk</i>	23
4.1	Representação visual do grafo $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$	32
4.2	Evolução da rede no tempo	32
4.3	Grafo agregado G	33
4.4	Centralidade dos vizinhos.	36
4.5	Utilização do valor da métrica dos vizinhos	36
4.6	Dois <i>snapshots</i> consecutivos ilustrando o aparecimento de uma ligação	38
4.7	Modelo para o estado de <i>link</i> dinâmico.	38
5.1	Cadeia de Markov utilizada no gerador.	45
6.1	Dinâmica considerando mobilidade	47
6.2	Algoritmos de difusão no agregado	48
6.3	Difusão orientada por centralidade - SopCast	49
6.4	Impacto do conhecimento do sistema (Δ) na proporção dos nós alcançados.	50
6.5	Difusão nos grafos sintéticos - Erdős e Rényi	52
6.6	Difusão nos grafos sintéticos - Barabási - Albert	52
6.7	Difusão nos grafos sintéticos - <i>Small World</i>	53
6.8	<i>Small World</i> - Baixa Dinâmica	54
6.9	<i>Small World</i> - Média Dinâmica	54
6.10	<i>Small World</i> - Alta Dinâmica	54
6.11	Número de réplicas geradas	57
6.12	Aceleração da difusão de informação	59

LISTA DE TABELAS

5.1	Parâmetros utilizados na geração das redes sintéticas.	44
5.2	Parâmetros utilizados pelo Gerador Markoviano	45
6.1	Correlação de Spearman	52
6.2	Valores do raio \mathcal{R} utilizados	56

SUMÁRIO

1	INTRODUÇÃO	12
1.1	CONTRIBUIÇÕES DO TRABALHO	14
1.2	ORGANIZAÇÃO DA DISSERTAÇÃO	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	MODELOS	16
2.2	MÉTRICAS DE CENTRALIDADE	18
2.3	ALGORITMOS DE DIFUSÃO EM REDES COMPLEXAS	19
2.3.1	Algoritmo de difusão Epidêmico	20
2.3.2	Algoritmo de difusão <i>Random Walk</i>	22
2.4	CONSIDERAÇÕES	23
3	TRABALHOS RELACIONADOS	25
3.1	VISÃO ESTÁTICA	25
3.2	VISÃO DINÂMICA	29
3.3	CONSIDERAÇÕES	30
4	ALGORITMOS DE DIFUSÃO BASEADOS EM MÉTRICAS DE CENTRALIDADE	31
4.1	MODELAGEM MATEMÁTICA	31
4.2	ALGORITMOS DE DIFUSÃO BASEADOS EM MÉTRICAS DE CENTRALIDADE	33
4.2.1	Custos dos Algoritmos de Difusão Utilizando Métricas de Centralidade.....	34
4.3	UTILIZAÇÃO DAS MÉTRICAS EM UM RAIOS DE ABRANGÊNCIA	35
4.4	MODELO DE PREVISÃO DE <i>LINKS</i>	37
4.5	MEDIDA DE INTERESSE	39
5	CONJUNTOS DE DADOS UTILIZADOS	41
5.1	<i>DARTMOUTH</i>	41
5.2	<i>INFOCOM</i>	42
5.3	<i>SOPCAST</i>	43

5.4	REDES SINTÉTICAS	43
5.5	MODELO <i>EDGE-MARKOVIAN</i>	44
6	RESULTADOS	46
6.1	ANÁLISE DOS MODELOS SOA E CUT	46
6.2	IMPACTO DA DINÂMICA NO PROCESSO DE DIFUSÃO	47
6.3	DIFUSÃO ORIENTADA POR CENTRALIDADE	48
6.4	REDES SINTÉTICAS	51
6.5	MODELO <i>EDGE-MARKOVIAN</i>	53
6.6	UTILIZAÇÃO DAS MÉTRICAS EM UM RAIÃO DE ABRANGÊNCIA	55
6.7	NÚMERO DE RÉPLICAS GERADAS NA REDE	57
6.8	PREVISÃO DE <i>LINKS</i>	58
6.9	CONSIDERAÇÕES	59
7	CONCLUSÃO	61
	REFERÊNCIAS	63

1 INTRODUÇÃO

Redes Complexas estão presentes em diversas áreas da vida cotidiana das pessoas (DOROGOVTSEV; MENDES, 2003; EASLEY; KLEINBERG, 2010; KOCAREV, 2010). Entender o comportamento e a formação topológica de diferentes redes complexas, por exemplo, redes de comunicação, sociais, biológicas e tecnológicas, tem atraído a atenção da comunidade de pesquisa ao longo dos últimos anos (BARBERA et al., 2011; HUI et al., 2011; KITSACK et al., 2010; MTIBAA et al., 2010; YUAN; MA, 2012).

Uma característica em comum entre estes sistemas é a dinamicidade. A estrutura física muda com o tempo, entidades são adicionadas ou retiradas e as ligações entre estas surgem ou desaparecem ao longo do período de observação (BOCCALETTI et al., 2006; ASH, 1997; WANG, 2002).

Diversas redes possuem um grau de dinamicidade elevado e, desta forma, suas características se modificam em um curto intervalo de tempo. Ao se analisar tais redes sem levar em consideração a sequência de modificações ocorridas no tempo, informações valiosas para uma análise detalhada do comportamento destes sistemas podem ser perdidas. Modelar e estudar tais sistemas com uma única fotografia que agrega uma grande quantidade de informação para um tempo longo de observação pode resultar em conclusões equivocadas sobre, por exemplo, o comportamento global e a interação entre as entidades que compõem estes sistemas (CARLEY, 2003; RANGANATHAN et al., 2002). Como consequência, não considerar a dinâmica pode comprometer propostas mais eficazes de gerência, construção e caracterização das redes encontradas na vida real (WATTS, 2003). Este fenômeno ocorre por se agregar, em uma única fotografia, um grande conjunto de modificações estruturais.

Até o momento, a maioria dos trabalhos presentes na literatura modelam redes complexas tendo como base uma visão estática da rede. Sendo assim, modelam o cenário estudado com uma única fotografia (BOGUñá et al., 2009; SIMSEK; JENSEN, 2008; KIM et al., 2002; ROSVALL et al., 2005).

A difusão de informação entre os elementos da rede, considerando a dinâmica do sistema, é um dos diversos problemas importantes que devem ser abordados no estudo de redes reais (ALVAREZ-HAMELIN et al., 2012; DEZSO, 2004; CARRINGTON et al.,

2005; VALENTE, 1996; NEWMAN, 2003; INTANAGONWIWAT et al., 2000; RANGO; MONTEVERDI, 2012). Difundir informação em uma rede pode ter diversos significados dependendo do sistema real em estudo. Por exemplo, em redes par-a-par (*P2P*), difundir informação significa distribuir arquivos no caso de sistemas *file sharing*; ou vídeo no caso de alguns sistemas de tempo real. No caso de redes biológicas, pode-se considerar a difusão de informação como sendo o processo de disseminação de uma doença ou de uma vacina entre pessoas que estabelecem algum contato direto ou indireto.

Claramente, o processo de difusão da informação em uma rede qualquer, seja uma rede social ou biológica, está intimamente atrelado ao comportamento e as características dos integrantes das mesmas (WANG et al., 2013). Por exemplo, em uma rede social, supõe-se que um indivíduo extrovertido tenha uma lista de contatos maior do que a lista de contatos de um indivíduo introvertido (UGANDER, 2013; WATANABE; SUZUMURA, 2013). Neste caso, se o objetivo é que uma informação alcance um maior número de indivíduos em um menor intervalo de tempo, indivíduos mais extrovertidos devem ser escolhidos para receber a informação a ser disseminada. Tais indivíduos possuirão uma lista maior de contatos e, desta forma, mais usuários poderão receber a informação em menor tempo. Este é um cenário simples onde a escolha da entidade que receberá a informação a ser propagada influencia o tempo de difusão de informação.

Ao ser introduzida a dinâmica na modelagem dos sistemas, quanto maior o número de informação disponível sobre a sequência de modificações ocorridas, mais provável será o encaminhamento da informação para o dispositivo mais propício a difundir a informação em menos tempo. Dentre as informações utilizadas na literatura para a escolha do melhor vizinho que receberá a informação estão o histórico cronológico dos contatos (LINDGREN et al., 2003), informações sobre padrões de mobilidade dos nós (TARIQ et al., 2006), informações de interações sociais entre os nós (HUI et al., 2011; MTIBAA et al., 2010) ou, o mais importante para este trabalho, informações sobre a centralidade dos usuários, que podem ser mensuradas por métricas definidas em teorias de grafos como grau, *betweenness*, *closeness*, dentre outras.

Devido a não-trivialidade na definição e modelagem da dinamicidade, a compreensão dos processos de difusão, ou cobertura da rede, se torna um desafio ainda maior se a dinâmica das redes complexas for considerada (ALVAREZ-HAMELIN et al., 2012). Como consequência, algoritmos de difusão propostos tendo como base cenários com característi-

cas dinâmicas, mas modelados como sistemas estáticos, podem apresentar vícios em seus resultados, devido ao fato da dinamicidade de tais sistemas não estar presente no modelo.

A grande maioria dos trabalhos não trata o processo de difusão observando a dinamicidade. Os modelos são estáticos e agrupam um grande período de monitoração (ADAMIC; ADAR, 2005; BOGUÑÁ et al., 2009; SIMSEK; JENSEN, 2008; KIM et al., 2002; ADAMIC et al., 2001; ROSVALL et al., 2005). Além disso, os trabalhos consideram apenas o percentual de nós que são alcançados pelos algoritmos, não se importando com o tempo necessário para realizar esta tarefa (BARBERA et al., 2011; KITSACK et al., 2010; NES; LATORA, 2008; ADAMIC et al., 2001). Esta pequena consideração da dinâmica de sistemas complexos ocorre principalmente devido à alta complexidade introduzida, tanto na modelagem quanto na análise das características comportamentais de maior relevância. Alguns trabalhos tratam da aceleração dos processos de difusão considerando a dinâmica, mas os algoritmos propostos necessitam de informações adicionais como as relações sociais (HUI et al., 2011; MTIBAA et al., 2010; RANGO; MONTEVERDI, 2012). Como consequência, estes algoritmos não necessariamente alcançam o desempenho que deveriam, pois utilizam informações que não obrigatoriamente estão disponíveis para todos os indivíduos.

O próprio conceito de dinâmica em um sistema real não é trivial e proposições devem ser estabelecidas *a priori* da análise a ser realizada. Neste trabalho consideram-se os seguintes fenômenos como fatores causadores da dinâmica. Surgimento e desaparecimento de arestas; mudanças ocorridas na lista de contatos dos nós em uma rede formada por uma aplicação para compartilhamento de vídeo; alteração do ponto de acesso em que um dispositivo está conectado em uma rede *wireless*; mudança da lista de dispositivos *Bluetooth* encontrados por participantes de uma conferência.

1.1 CONTRIBUIÇÕES DO TRABALHO

Este trabalho propõe e analisa algoritmos de difusão de informação em redes complexas, considerando a dinâmica do sistema. Será utilizada uma modelagem de grafos onde a dinâmica do sistema é considerada, buscando com isso uma maior correlação do modelo com a realidade. Nesse sentido, as contribuições deste trabalho são:

- i Algoritmos tradicionais utilizados no processo de difusão de informação, ao ignorar a

- dinâmica do sistema, subestimam o tempo necessário para a informação atingir todos os nós (ou seja, o tempo de cobertura).
- ii Proposta e análise de diferentes algoritmos de difusão em redes dinâmicas, buscando diminuir o tempo necessário para distribuir informação em tais redes. Estes algoritmos consideram a importância das entidades da rede, utilizando para isto as características de centralidade dos indivíduos participantes.
 - iii Aplicação de tais algoritmos em redes reais e sintéticas.
 - iv Variação simples do algoritmo de difusão proposto, melhorando o tempo de difusão de informação. Tal variação baseia-se na centralidade dos vizinhos pertencentes a um raio de abrangência.
 - v Desenvolvimento de um método de previsão de arestas, onde um determinado nó pode prever a formação topológica da rede nos instantes subsequentes, utilizando assim tal informação para acelerar o processo de difusão.

Difusão de informação orientada por *bias* não é uma característica recente (LINDGREN et al., 2003; HUI et al., 2011; KITSACK et al., 2010; BARBERA et al., 2011). Porém, neste trabalho são realizadas comparações dos resultados obtidos com diversas métricas de centralidade, quando aplicadas na difusão de informação em redes complexas dinâmicas.

1.2 ORGANIZAÇÃO DA DISSERTAÇÃO

Este trabalho está organizado da seguinte maneira. Os conceitos teóricos que fundamentam a análise, incluindo a dinâmica de redes de comunicação, são apresentados no Capítulo 2. O Capítulo 3 apresenta os trabalhos relacionados a esta dissertação. No Capítulo 4 são apresentados algoritmos de difusão que consideram métricas topológicas de centralidade para acelerar a difusão de informação na rede. As redes de comunicação analisadas neste trabalho são descritos no Capítulo 5. Os resultados numéricos alcançados neste trabalho estão apresentados no Capítulo 6. As considerações finais deste trabalho estão apresentadas no Capítulo 7.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os fundamentos teóricos utilizados neste trabalho. Serão apresentadas definições matemáticas utilizadas na modelagem dos sistemas, os algoritmos básicos de difusão de informação presentes na literatura, além das métricas topológicas que são utilizadas nos algoritmos de difusão propostos neste trabalho.

2.1 MODELOS

Uma rede complexa pode ser matematicamente representada por um grafo. Seja $G(V, E)$ a representação desta rede, sendo V o conjunto de nós e E o conjunto de arestas. Abaixo são apresentados os principais modelos clássicos de redes complexas, encontrados na literatura.

Grafos Aleatórios: O estudo sistemático de grafos aleatórios foi iniciado por Erdős e Rényi em 1959, com o propósito original de estudar, por meio de métodos probabilísticos, as propriedades dos grafos à medida que se aumentava o número de ligações aleatórias. O termo “grafo aleatório” refere-se à natureza desordenada do arranjo das ligações entre diferentes nós (BOCCALETTI et al., 2006).

Em seu artigo, Erdős e Rényi propuseram um modelo para gerar grafos aleatórios com N nós e K ligações. Iniciando com N nós desconectados, o grafo aleatório é gerado por pares selecionados aleatoriamente, até que o número de arestas seja igual a K (ERDÖS; RÉNYI, 1959).

Small World: Os grafos que contêm a propriedade *Small World* possuem elevado coeficiente de agrupamento e uma pequena distância média entre os nós (WATTS; STROGATZ, 1998). O modelo Watts e Strogatz é um método para se construir grafos possuindo a propriedade *Small World*. O ponto de partida da geração de grafos é um anel com N nós, onde cada nó é simetricamente ligado a M nós mais próximos no grafo, somando um total de arestas $K = \frac{N*M}{2}$. A partir desta formação, para cada nó, cada ligação deste nó com seu vizinho no sentido horário é retirada com probabilidade p e reconectada a um nó aleatório (WATTS; STROGATZ, 1998).

A Figura 2.1 mostra que, se $p = 0$, o grau dos nós do grafo gerado é totalmente regular, enquanto que, se $p = 1$, o modelo produz um grafo aleatório. Para valores intermediários de p , o modelo gera um grafo com a propriedade *Small World*, com um coeficiente de clusterização não trivial. A Figura 2.1 mostra o impacto do parâmetro p na criação do grafo pelo modelo Watts e Strogatz.

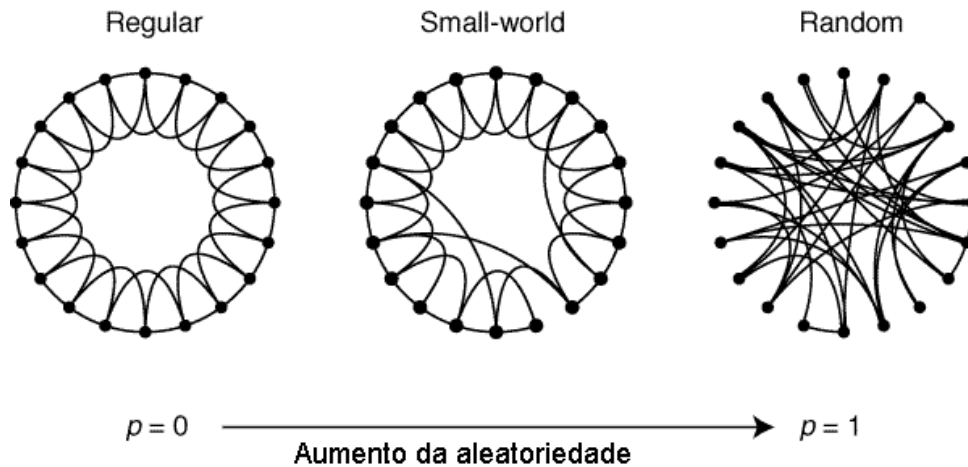


Figura 2.1: Influência do parâmetro p no modelo *Small World*. Figura extraída de Watts e Strogatz (1998).

Scale-Free: O modelo das redes de escala livre (*scale-free network*) foi proposto pelo cientista Albert-László Barabási na década de 1990. Este modelo foi criado em oposição ao modelo dos grafos aleatórios, proposto por Erdős e Rényi e ao modelo *Small World*, proposto por Watts e Strogatz. Segundo este modelo, para a geração do grafo, parte-se de um pequeno número de nós e, em cada instante de tempo, cria-se um novo nó que é ligado a os demais nós existentes.

O modelo Barabási - Albert baseia-se em dois componentes básicos: crescimento e conexão preferencial (BARABÁSI; ALBERT, 1999). Neste modelo a probabilidade de um novo vértice se conectar a um já existente depende da conectividade deste último. Em outras palavras, quanto maior o número de conexões de um vértice, maior a probabilidade de um novo vértice se conectar a este.

Ao iniciar com m_0 nós isolados, a cada passo de tempo t , onde $t \in [0, N - m_0]$, ou seja, $t = 0, 1, 2, 3 \dots N - m_0$, um novo nó j é adicionado no grafo com $m \leq m_0$ conexões (BARABÁSI; ALBERT, 1999).

A Figura 2.2 apresenta as diferenças entre o modelo dos grafos aleatórios (a) e o modelo das redes de escala livre (b), considerando a distribuição do grau dos nós. No primeiro modelo, o grau da maioria dos nós está próximo do grau médio. No segundo modelo existem alguns poucos nós com um número de conexões em uma escala bem maior que os demais, formando assim um gráfico que segue a lei de potência.

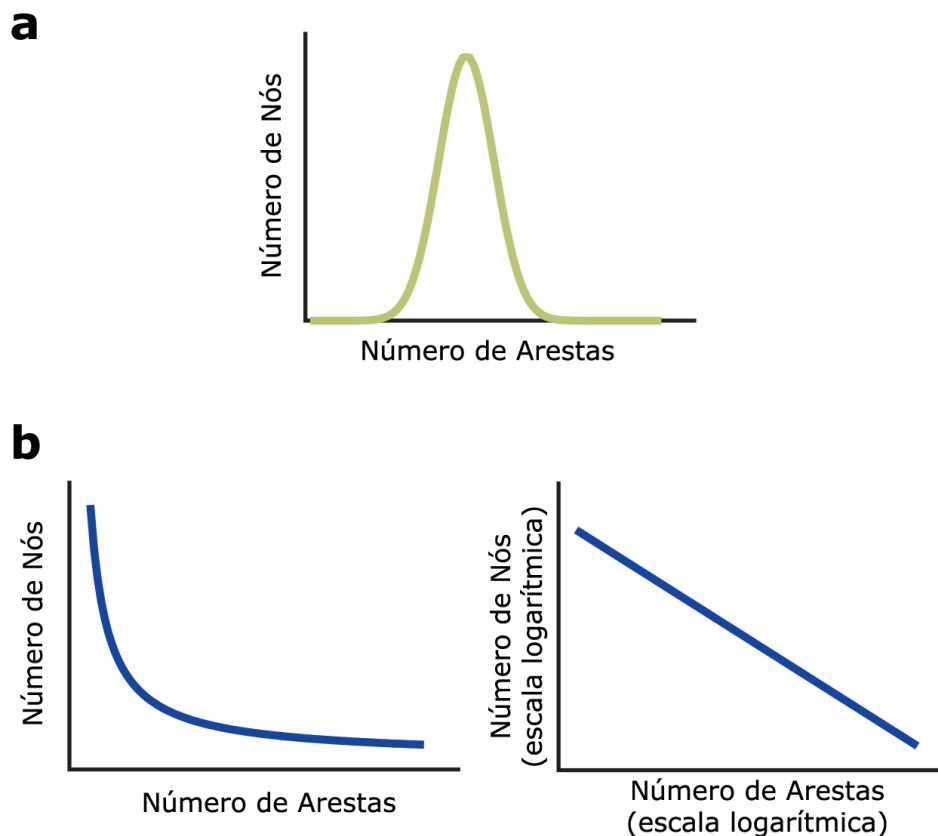


Figura 2.2: Diferença entre o modelo de grafos aleatórios e o modelo *scale-free*. Figura extraída de Barabási e Bonabeau (2003).

2.2 MÉTRICAS DE CENTRALIDADE

Métricas topológicas são definidas como medidas baseadas em atributos estruturais de um grafo. Estas métricas podem considerar cada nó individualmente, ou então o grafo como um todo. Neste trabalho são utilizadas as métricas relacionadas à importância dos nós, conhecidas como métricas de centralidade. Tais métricas serão combinadas com algoritmos de difusão para acelerar o processo de difusão de informação em redes dinâmicas.

As métricas consideradas neste trabalho são:

Grau: O grau de um nó v é definido como o total de arestas incidentes a este nó. Este valor é normalizado, dividindo o número de arestas incidentes pelo número de nós na rede. Sendo assim, o valor da métrica para um nó v será a fração do grau, $d(v)$, dividido pelo valor máximo do grau. A métrica está definida da seguinte forma:

$$Grau(v) = \frac{d(v)}{\max_{\forall u \in V} d(u)}. \quad (2.1)$$

Betweenness: O *Betweenness* de um nó é a fração dos caminhos mínimos que ligam qualquer par de nós e que passam por ele. Em outras palavras, seja $\sigma_{u,j}$ o número de caminhos mínimos entre u e j , e $\sigma_{u,j}(v)$ o número total de caminhos mínimos entre u e j que passam por v . A métrica *Betweenness* é definida como:

$$Betweenness(v) = \sum_{\forall v \neq u \neq j \in V} \frac{\sigma_{u,j}(v)}{\sigma_{u,j}}. \quad (2.2)$$

Closeness: A métrica *Closeness* de um nó captura o quão perto este nó está de todos os nós que podem ser alcançados a partir deste na rede. Dado o tamanho de um caminho mínimo entre v e j , definido por $l(v, j)$, a métrica *Closeness* é dada por:

$$Closeness(v) = \left[\sum_{\forall v \neq j, j \in V} l(v, j) \right]^{-1}. \quad (2.3)$$

2.3 ALGORITMOS DE DIFUSÃO EM REDES COMPLEXAS

Suponha que o nó u necessite enviar uma informação ao nó v . Em cada instante de tempo, cada nó pode armazenar as mensagens que tem ou então encaminhá-las para qualquer um de seus vizinhos.

Um caminho $u \rightarrow v$ é criado ao longo do tempo, mesmo que não existam conexões entre u e v em um dado instante de tempo (*snapshot*). O caminho é construído sobre o

grafo da rede em função de sua dinâmica, até que a informação chegue ao nó destino. É possível notar, porém, que mesmo se não houver algum caminho direto entre u e v em um dado instante de tempo, um caminho pode existir entre estes nós, na medida em que a rede se altere devido à dinâmica de sua topologia.

Conforme apresentado em Basu et al. (2010), os modelos *Store-or-Advance* (SoA) e *Cut-Through* (CuT) definem o conjunto de entidades alcançáveis de cada um dos membros do sistema, conjunto este tratado neste trabalho como “vizinhos”. De acordo com Basu et al. (2010), no modelo SoA, um nó v qualquer pode encaminhar a informação apenas aos seus vizinhos diretos, mesmo que arestas entre os vizinhos dos vizinhos estejam ativas no instante de tempo analisado. O modelo SoA é equivalente ao modelo *store-carry-forward*, apresentado em Acer et al. (2011). No modelo CuT, um nó pode enviar a mensagem para qualquer outro nó em sua componente ligada. Sendo assim, toda esta componente pode ser percorrida instantaneamente ou, pelo menos, em uma escala de tempo muito menor do que a dinâmica do sistema.

A Figura 2.3 mostra graficamente a diferença entre os modelos SoA e CuT. Conforme indicado na figura, para um o instante de tempo t_9 , o modelo de difusão SoA alcança seis nós. Em contrapartida, para o mesmo intervalo de tempo, o modelo de difusão CuT alcança nove nós. A maior velocidade na difusão do conteúdo deve-se ao fato do modelo CuT permitir que sejam alcançados todos os nós pertencentes a componente ligada no instante de tempo que está sendo analisado.

A seguir são descritos os dois algoritmos base de difusão de informação: Epidêmico e *Random Walk*.

2.3.1 ALGORITMO DE DIFUSÃO EPIDÊMICO

Conforme apresentado em Demers et al. (1987), Vogels et al. (2003) e Vahdat e Becker (2000), a denotação *Epidêmico*, adotada nesse trabalho, refere-se à implementação onde todos os vizinhos habilitados recebem a informação a cada iteração. A recepção da informação é independente do modelo CuT ou SoA adotado. Caso seja o modelo SoA, somente os vizinhos diretos receberão a informação. Caso seja considerado o modelo CuT, todos as entidades pertencentes a componente ligada receberão a mensagem. A Figura 2.4

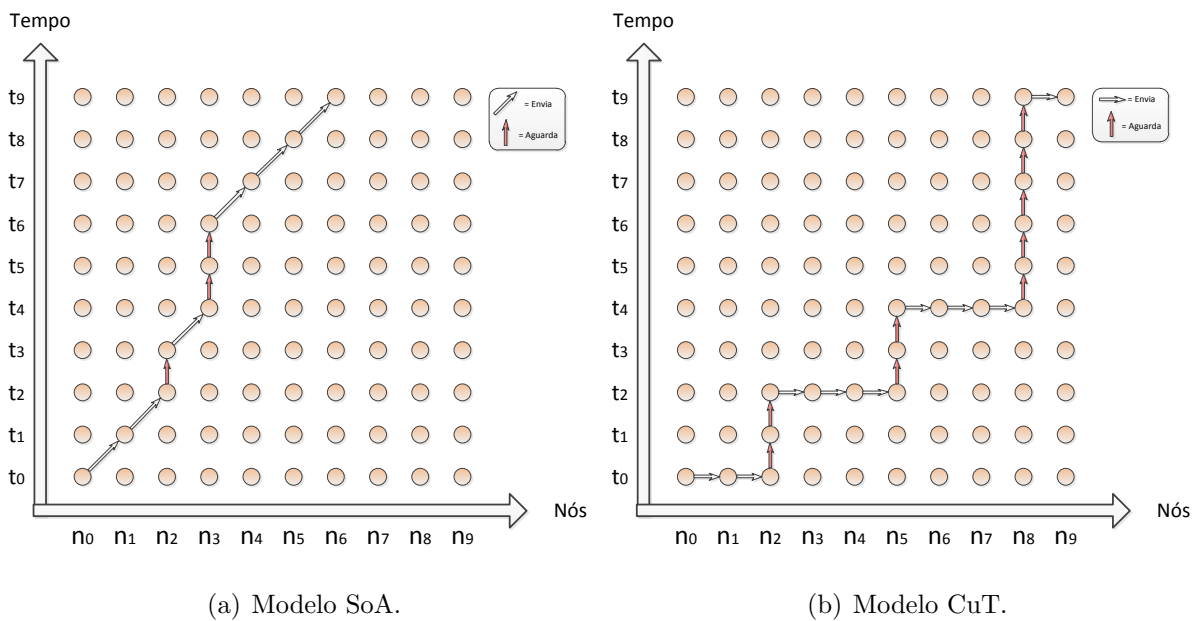


Figura 2.3: Diferença entre os Modelos de SoA e CuT.

apresenta o impacto dos modelos SoA e CuT no Epidêmico. No modelo SoA, somente os vizinhos diretos 2 e 3 recebem a informação no instante inicial da difusão. Já no modelo CuT, todas as entidades pertencentes a componente conexa recebem a mensagem.

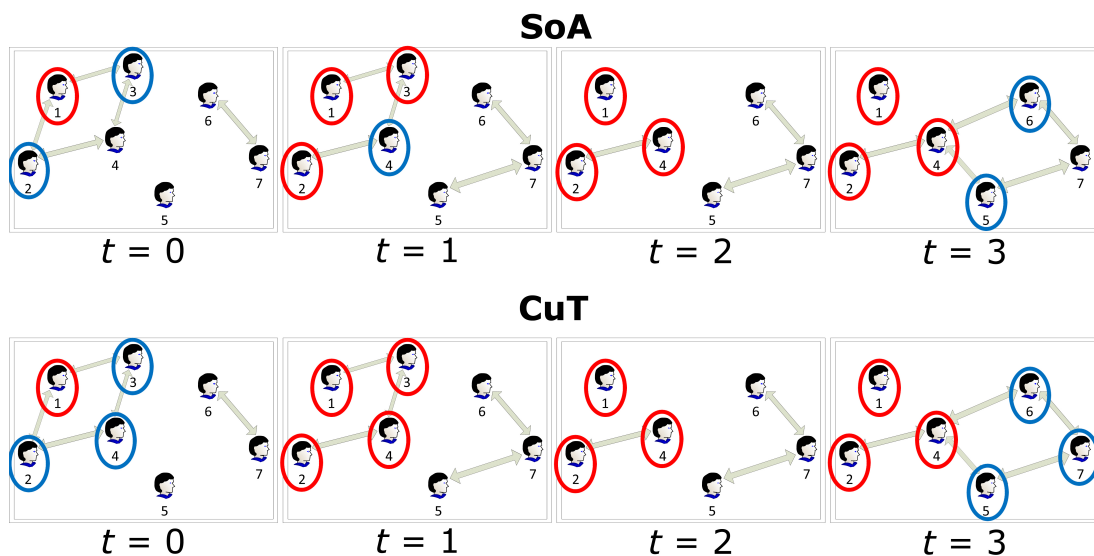


Figura 2.4: Impacto dos modelos SoA e CuT no algoritmo *Epidêmico*.

É fácil notar que este modelo dará o menor tempo de difusão partindo de uma origem, tendo em vista que todos os vizinhos dos nós que possuem a informação receberão o

dado qualquer que seja o instante de tempo analisado. Isto porque, qualquer que seja o caminho escolhido para envio da informação pelos outros métodos, este caminho estará contido no conjunto de caminhos gerados pelo algoritmo Epidêmico. No entanto, o custo computacional deste método, baseado no volume do tráfego gerado e também no número total de mensagens trocadas entre os nós, é elevado. Como consequência, a implementação deste algoritmo em sistemas reais é complexa e pode-se tornar inviável. Por exemplo, em um sistema par a par para compartilhamento de vídeo, ao utilizar tal abordagem será gerada uma sobrecarga do sistema como um todo, o que prejudicaria a entrega do vídeo aos dispositivos da rede.

A complexidade deste algoritmo baseada no volume de tráfego gerado e também no número de mensagens trocadas é da ordem de $O(|V|^2)$.

2.3.2 ALGORITMO DE DIFUSÃO *RANDOM WALK*

Conforme descrito em Lovasz (1993), o funcionamento do algoritmo de difusão *Random Walk* é simples: dado um grafo qualquer e um nó aleatório v como ponto de partida, um vizinho u de v é escolhido aleatoriamente e a informação é repassada a este. De forma análoga, um vizinho i escolhido aleatoriamente entre os vizinhos de u , recebe a informação na próxima iteração do algoritmo. A sequência aleatória dos nós selecionados neste caminho é definida como um passeio aleatório no grafo, ou seja, um *Random Walk*.

Dada o comportamento do algoritmo *Random Walk*, é razoável que o tempo de difusão da informação seja maior que nos algoritmos epidêmicos. Adicionalmente, este algoritmo não considera características do sistema real que podem ser usadas para acelerar a difusão da informação entre as entidades da rede. No caso de sistemas onde o tempo de entrega da informação é crucial, torna-se importante identificar nós que potencialmente possam acelerar o processo de difusão (HUI et al., 2011; YUAN; MA, 2012; MTIBAA et al., 2010).

Nos resultados apresentados na Seção 6, são consideradas duas implementações diferentes, tanto do algoritmo *Random Walk* quanto do Epidêmico: a tradicional, como descrita anteriormente, que pode ser vista como um caso particular do modelo SoA, onde a cada instante de tempo somente os vizinhos diretos podem ser escolhidos para receber a informação; e a implementação baseada no modelo CuT, onde considera-se que cada um dos nós dentro de uma componente ligada, para um instante de tempo qualquer, poderá ser escolhido para receber a informação difundida. A Figura 2.5 apresenta o impacto dos

modelos SoA e CuT no algoritmo *Random Walk*. No modelo SoA, o algoritmo seleciona um nó aleatoriamente entre seus vizinhos diretos. Já no modelo CuT, qualquer nó dentro da componente ligada pode ser escolhido pelo algoritmo.

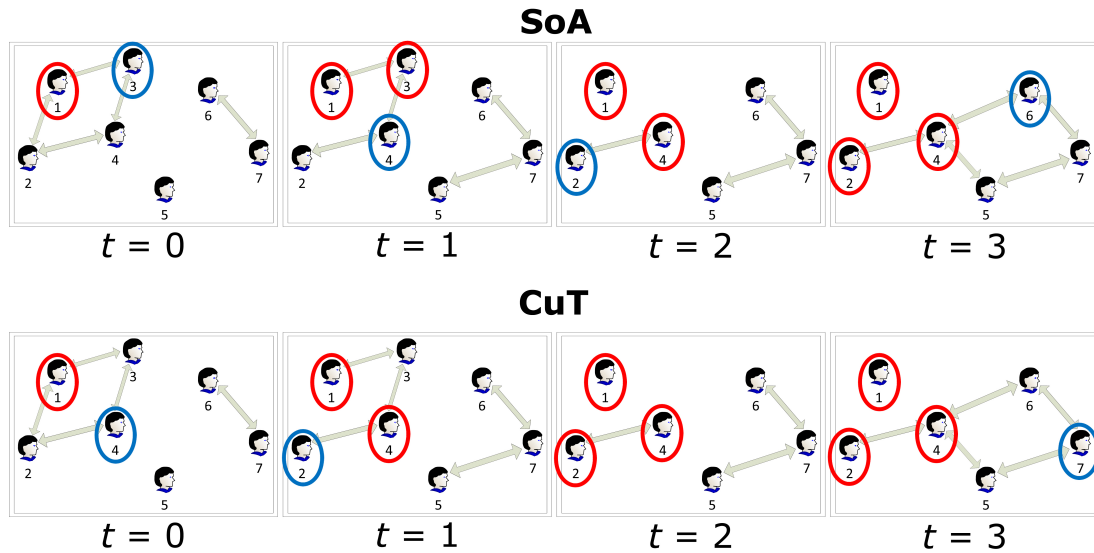


Figura 2.5: Impacto dos modelos SoA e CuT no algoritmo *Random Walk*.

A complexidade deste algoritmo baseada no volume de tráfego gerado e também no número de mensagens trocadas é da ordem de $O(|V|)$.

É fácil notar que este método dará um volume de tráfego menor e também um número menor de réplicas que o algoritmo Epidêmico apresentado na Seção 2.3.1, porém o tempo de difusão da informação na rede deste algoritmo é superior, conforme será mostrado no Capítulo 6. Cabe ainda ressaltar que, dado por um efeito em cascata, o número de mensagens trocadas no algoritmo Epidêmico tem um fator de crescimento muito maior que o algoritmo *Random Walk*.

2.4 CONSIDERAÇÕES

Neste capítulo foi apresentado o fundamento teórico necessário ao desenvolvimento desta dissertação. Foram apresentados os algoritmos básicos de difusão de informação presentes na literatura, Epidêmico e *Random Walk*, além das métricas topológicas utilizadas nos algoritmos de difusão propostos neste trabalho. Foi apresentada a formalização dos custos destes algoritmos básicos, onde foi possível concluir que o algoritmo Epidêmico

dará o menor tempo de difusão, porém com um custo elevado, baseado no volume de tráfego gerado. Já o algoritmo *Random Walk* possui um baixo custo computacional, porém o tempo de difusão será maior que os algoritmos epidêmicos.

3 TRABALHOS RELACIONADOS

É grande o interesse da comunidade científica em modelar e estudar diversos tipos de redes que fazem parte do nosso dia a dia. No entanto, considerar a dinâmica destes sistemas no processo de modelagem não é algo trivial. Da mesma forma, o interesse em difundir informação entre entidades destas redes ganhou importância nos últimos anos, principalmente em áreas como divulgação de produtos comerciais e até mesmo controle de infecção de doenças (NEWMAN, 2010).

Neste capítulo serão apresentados os principais trabalhos presentes na literatura onde é estudado o problema de difundir informação em redes complexas.

3.1 VISÃO ESTÁTICA

Nesta seção serão apresentados os trabalhos que tratam da difusão de informação em sistemas onde a dinâmica é compactada em uma única visão. Em outras palavras, não são consideradas mudanças sucessivas na relação entre as entidades do sistema e os modelos de tais sistemas são baseados em um único grafo.

As redes formadas por pessoas em sua vida cotidiana possuem uma infinidade de interações, por meio do qual as pessoas se encontram, fazendo com que ideias se espalhem e doenças infecciosas propagam-se dentro de uma sociedade (CALDARELLI; VESPIGNANI, 2007; JE, 1992). Diante disso, identificar os membros mais “difusores” de informação é um passo importante para a otimização da utilização dos recursos disponíveis e assegurar uma disseminação mais eficiente de informações.

O trabalho proposto em Kitsak et al. (2010) discute como a organização topológica da rede pode ser utilizada no auxílio do processo de disseminação de informação. Em Kitsak et al. (2010) é mostrado que nós com um grau elevado, mas presentes na periferia da rede tem pouca influência no tempo de propagação de informação na rede. Em contrapartida, nós com um grau não tão elevado, mas pertencente ao núcleo da rede possuem uma grande importância na disseminação de informação.

Outra solução para o problema de difundir informação em redes presentes no dia a dia das pessoas foi proposta em Hui et al. (2011). Neste trabalho, os autores buscam explorar as características da mobilidade humana para realizar a cobertura da rede em termos

de distribuição de informação. O trabalho propõe um algoritmo que se baseia em duas características principais dos usuários da rede: centralidade e comunidade. O trabalho parte da premissa que, dentro de uma mesma comunidade de pessoas, os indivíduos que são considerados mais populares possuem uma interação maior que os demais, propiciando assim um contato com todos os outros indivíduos da comunidade.

O algoritmo proposto em Hui et al. (2011) é denominado *BUBBLE Rap* e é especificado da forma que segue. O indivíduo que deseja repassar uma mensagem a outro indivíduo encaminha a mensagem para um vizinho que, dentre os seus possíveis contatos, possui uma centralidade maior dentro de toda a sociedade. Cada indivíduo que recebe a mensagem continua fazendo este mesmo processo, até que a mensagem seja entregue a um indivíduo que pertença a mesma comunidade do destino da informação. A partir deste ponto, é feito um ranqueamento da centralidade local, identificando a centralidade de cada um dos indivíduos que pertencem à comunidade do destinatário, fazendo com que a mensagem seja encaminhada àquele que possui a maior centralidade local. O algoritmo continua encaminhando a mensagem se orientando agora pela centralidade local dos indivíduos, calculada dentro da comunidade do destinatário. O processo segue desta forma até que a mensagem seja entregue ao destinatário ou então se perca.

De acordo com os autores, o algoritmo segue a intuição dos seres humanos. Quando se deseja repassar uma informação a uma determinada pessoa, primeiramente a mensagem é repassada às pessoas mais populares entre os contatos do indivíduo que deseja difundir uma informação. Quando a mensagem chega à comunidade da pessoa de destino, a primeira pessoa que recebe esta informação dentro desta comunidade encaminha para o mais popular dentre os conhecidos da comunidade. O processo continua seguindo esta característica, até que a mensagem chegue ao destino ou então deixe de ter relevância.

Porém, em Hui et al. (2011), os autores não levam em consideração as alterações das centralidades e das comunidades ao longo do tempo. No mundo real, estas características certamente não são estáticas, conforme mostrado em Ferreira et al. (2013). Desconsiderar as alterações sofridas pela sociedade ao longo do tempo na modelagem do sistema pode levar a cálculos equivocados do tempo necessário para entregar uma mensagem ao seu devido destino.

Em Lindgren et al. (2003) é tratado o problema de roteamento em redes que se alteram ao longo do tempo seguindo um padrão de movimentação definido. Em tais redes não há

garantia de que exista um caminho totalmente conexo entre a origem e o destino a todo o momento, o que torna os protocolos tradicionais de roteamento incapazes de entregar as mensagens que são trocadas na rede.

Assim, é proposto em Lindgren et al. (2003) o algoritmo *PROPHET*, que se baseia nos padrões de mobilidade dos indivíduos. Os autores partem do intuito de que usuários reais não se movimentam de forma aleatória, mas seguindo certos padrões, tornando a movimentação previsível. Sendo assim, estas informações podem ser usadas para difundir a informação na rede. Para cada par de nós, o algoritmo define uma probabilidade p chamada *delivery predictability* que indica a chance de que estes nós consigam trocar mensagem na rede com suas movimentações. Esta probabilidade é utilizada na escolha do encaminhamento da informação pela origem, escolhendo o nó que possua uma maior probabilidade de entrega da informação ao seu destino.

O algoritmo *PROPHET* requer um conhecimento prévio da rede a ser utilizada, para que se chegue a o valor da probabilidade p . Dessa forma, ficaria inviável a aplicação de tal algoritmo em redes em que o padrão de evolução e movimentação é desconhecido. Além disso, a dinâmica da rede não é explorada de forma eficiente. É tratada a mobilidade dos usuários, mas não é considerado o instante de tempo em que as mobilidades ocorrem, inviabilizando assim uma definição da ordem dos acontecimentos.

Em Barbera et al. (2011), os autores tratam do problema do congestionamento das redes de celulares, gerada pelo volume de transmissão de dados em regiões metropolitanas. O objetivo principal é cobrir o maior percentual possível de nós integrantes no sistema, canalizando o tráfego a um conjunto de usuários importantes, denominados *VIPs*, diminuindo assim o número de réplicas das mensagens na rede. O tráfego canalizado seria destinado principalmente a aplicações que requerem troca de grande quantidade de dados.

A solução proposta pelos autores é denominada *VIP Delegation*, onde alguns usuários da rede de telefonia são escolhidos de acordo com certas características baseadas em aspectos sociais e, com sua movimentação, são capazes de se comunicarem todos os demais usuários da rede. Nesse trabalho os autores classificam os usuários *VIPs* de acordo com os valores de suas métricas de centralidade de *Betweenness* (FREEMAN, 1977), *Closeness* (OKAMOTO et al., 2008) e grau, utilizando também a métrica *Pagerank* (BRIN; PAGE, 1998).

A rede formada pelos portadores dos *smartphones* é observada durante um período

de sete dias e, a partir desta observação, é gerado um grafo agregando a informação de todo o período de observação do sistema. Com este grafo, os *VIPs* são identificados e então destinados a distribuir toda a informação a todos os demais dispositivos da rede, com contatos diretos, gerados a partir de sua movimentação.

Neste trabalho proposto por Barbera et al. (2011), os autores exploraram somente a formação de parcerias dos usuários *VIPs*, provida pela mobilidade, com todos os destinos, visando a entrega de toda a informação aos demais usuários da rede. Porém, é gerado um grafo final que agrega todas as parcerias formadas ao longo do tempo, perdendo assim a noção de dinâmica do sistema e a ordem cronológica dos acontecimentos. O trabalho poderia ter utilizado ainda uma abordagem *multi-hop* (GERLA et al., 1999), fazendo com que a informação fosse repassada não apenas pelos usuários *VIPs*, mas também por outros membros da rede. Dessa forma, esta abordagem continuaria auxiliando na diminuição das requisições de informação aos provedores, além também de prover uma diminuição no tempo de propagação da mensagem por toda a rede.

Em Yuan e Ma (2012) os autores propõem uma estratégia que elege um dispositivo integrante da rede como sendo o *Hug* (*Human Gathering Point*). Este dispositivo é um integrante que, com certa probabilidade, fará a ponte entre a origem e o destino. Diante disso, o roteamento é dividido em duas fases. Na primeira, denominada *quick spray phase*, o dispositivo de origem da informação encaminha a mensagem para todos os vizinhos, até que, em um dado instante, a informação chega ao dispositivo *Hug*. A partir deste ponto, se inicia a fase *slow spray phase*. Nesta fase, o dispositivo *Hug* encaminha a mensagem a um número L de vizinhos a cada instante de tempo. Na medida em que o tempo passa, o valor de L decai linearmente.

Esta proposta também não considera as alterações que a rede sofre com o passar do tempo, fazendo com que a rede seja representada por apenas um único grafo, com toda a informação agregada. Além do mais, os resultados apresentados por Yuan e Ma (2012) ficam bastante dependentes dos bons resultados do método de escolha do dispositivo *Hug*. Sendo assim, uma escolha ruim do dispositivo *Hug* continuaria gerando grande replicação da informação, além da mensagem não alcançar o destino em tempo hábil.

3.2 VISÃO DINÂMICA

Nesta seção serão apresentados os principais trabalhos que tratam da difusão de informação em redes complexas e que modelam a rede levando em consideração a dinâmica do sistema.

O trabalho apresentado em Mtibaa et al. (2010) propõe um algoritmo denominado *PeopleRank*, onde os membros integrantes da rede são classificados por uma média ponderada de informações sociais. Nesta proposta, o algoritmo proporciona um maior peso aos dispositivos pertencentes a pessoas socialmente conectadas com usuários providos de alto nível de popularidade nas redes sociais.

O algoritmo oportunista *PeopleRank* classifica a “importância” de um usuário no grafo social. Este ranqueamento apresentado em *PeopleRank* é inspirado no *PageRank*, um algoritmo que mede o quão acessível é uma página na *Web* partindo das demais, algoritmo este utilizado pelo motor de busca do *Google* (GOOGLE, 2013) (BRIN; PAGE, 1998). Análogo ao *PageRank*, o *PeopleRank* identifica os usuários mais populares no contexto social capazes de realizar a cobertura dos outros usuários da rede. A diferença crucial entre os dois algoritmos é que o *PageRank* é centralizado e o *PeopleRank* é distribuído.

De acordo com os autores, com o surgimento de plataformas de redes sociais como o *MySpace* (MYSPACE, 2013), *Orkut* (ORKUT, 2013) e *Facebook* (FACEBOOK, 2013), informações sobre a interação social dos usuários tornou-se rapidamente disponível. Além disso, as informações sobre as parcerias formadas na rede de dispositivos são inconstantes. Em contrapartida, ainda segundo os autores, as informações sociais são bastante estáveis. Sendo assim, o algoritmo *PeopleRank* utiliza informações sociais dos usuários a fim de fornecer dados suficientes para a realização do roteamento oportunista.

O algoritmo *PeopleRank*, ao utilizar informações de perfis presentes nas redes sociais dos usuários, parte da premissa de que os usuários presentes na rede de dispositivos móveis tenham perfis em tais redes sociais, fato este que não é uma tautologia, haja vista que não existe o pré-requisito de se ter um perfil em redes sociais para ser um usuário de *smartphone*, além ainda da necessidade de que a popularidade presente nestes perfis deva corresponder à realidade dos membros integrantes da rede. Nesta dissertação, são utilizadas apenas características topológicas, a fim de classificar os nós mais importantes para a disseminação da informação.

3.3 CONSIDERAÇÕES

Neste capítulo foram apresentados os trabalhos atuais mais relevantes que tratam da modelagem de redes presentes na vida real, além das diversas soluções presentes na literatura que tratam do problema de difusão de informação em redes que se modificam no tempo.

Como descrito, a maioria das propostas não considera a dinâmica dos sistemas. Além do mais, quando os trabalhos reconhecem a dinâmica da rede, esta não é modelada por completo. Foram apresentadas também algumas propostas presentes na literatura que necessitam de informações que eventualmente não estão disponíveis, ou até mesmo não existam para todos os membros da rede.

4 ALGORITMOS DE DIFUSÃO BASEADOS EM MÉTRICAS DE CENTRALIDADE

Neste capítulo serão apresentados os algoritmos de difusão propostos nesta dissertação. Estes algoritmos visam minimizar o tempo gasto no processo de difusão da informação, minimizando também o volume de tráfego e o número de réplicas de mensagens na rede.

4.1 MODELAGEM MATEMÁTICA

Uma rede será representada matematicamente por um grafo. Seja $G(V, E)$ a representação desta rede, sendo V o conjunto de nós e E o conjunto de arestas. Este conjunto de nós V pode representar pessoas querendo se comunicar, dispositivos eletrônicos em busca de troca de dados e até mesmo indivíduos enfermos contaminando todo o resto da população.

Seja o período total de observação do sistema real denotado por T_N , iniciado no tempo $T_1 = 0$. A modelagem estática da rede será definida pelo grafo estático G , agregando neste todas as informações capturadas na rede no período de observação T_1 a T_N . Como um dos objetivos desta dissertação é verificar o impacto da dinâmica dos sistemas no processo de difusão de informação, a seguir é definido o modelo matemático que representa esta dinâmica.

Seja $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$ a representação dinâmica do grafo agregado G , sendo \mathcal{V}_t e \mathcal{E}_t o conjunto nós e arestas existentes no instante t , respectivamente. Sendo assim, o grafo \mathcal{G}_t e os conjuntos \mathcal{V} e \mathcal{E} estão indexados no tempo t , com $T_1 \leq t \leq T_N$. A análise do comportamento dinâmico da rede é feita através de uma sequência de grafos organizados em fotografias (*snapshots*) incrementais no tempo. A Figura 4.1 exemplifica a construção de \mathcal{G}_t . Desta forma, busca-se manter a sequência evolutiva da rede, fazendo com que o evento e_1 , ocorrido na rede no instante t_1 , e o evento e_2 ocorrido no instante t_2 continuem sendo representados em instantes diferentes na modelagem da rede.

Para exemplificar a diferença entre os modelos apresentados, considere uma rede com dispositivos móveis, onde os seus usuários se encontram em diferentes instantes de tempo.

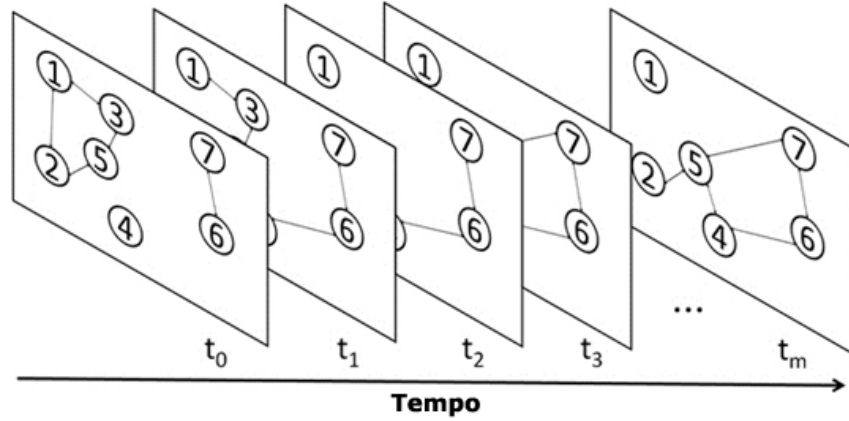


Figura 4.1: Representação visual do grafo $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$.

A Figura 4.2 apresenta um exemplo da rede hipotética de dispositivos móveis em três instantes de tempo diferentes. É possível notar que existem conexões ocorrendo no mesmo instante de tempo e também em instantes diferentes. Neste contexto, é desejável que a modelagem do sistema preserve esta característica. A Figura 4.3 apresenta o grafo agregado G desta mesma rede hipotética. Este grafo não preserva a ordem temporal das conexões da rede, mantendo somente um *snapshot* de todo o período de observação. Ao modelar um sistema real é desejável que a ordem dos acontecimentos seja preservada. Ao ignorar tal informação na modelagem, os algoritmos de difusão de informação baseados neste modelo podem apresentar resultados que não correspondem a realidade, conforme discutido no Capítulo 3.

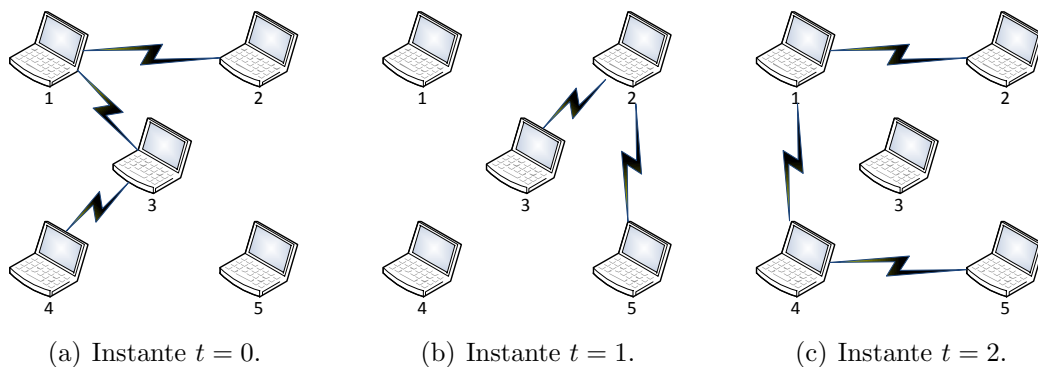


Figura 4.2: Evolução da rede no tempo.

A duração de cada *snapshot* $g_i \in \mathcal{G}_t$ é definida pelo valor de Δ , de forma que as arestas presentes em g_i serão aquelas que ocorrerem no intervalo de tempo $[T_{i*\Delta}, T_{(i+1)*\Delta})$; onde

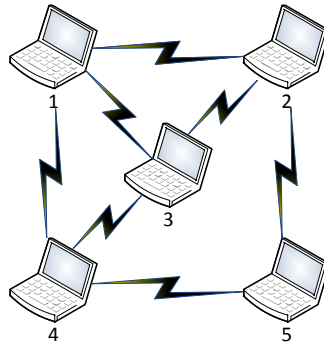


Figura 4.3: Grafo agregado G da rede anterior.

$0 \leq i \leq N$. Assim, para cada rede analisada, serão gerados $|\mathcal{G}_t| = \lceil T_N/\Delta \rceil$ *snapshots*. Esta representação permite capturar a evolução da rede no espaço e no tempo, ou seja, mudanças estruturais são capturadas a cada Δ unidades de tempo. Claramente, se $\Delta = T_N$, é representada uma única fotografia do sistema, desconsiderando a dinâmica, reduzindo a análise ao grafo estático G . Em contrapartida, a sequência de grafos \mathcal{G}_t pode ser interpretada como um conjunto de fotografias do sistema, cada uma delas representando as mudanças estruturais ocorridas em um *snapshot* em particular.

A escolha do parâmetro Δ influencia na dinamicidade capturada na modelagem do sistema. Quanto menor o seu valor, maior é a aproximação de uma análise dinâmica formada por várias fotografias. Para obter uma análise com menor granularidade, Δ pode ser definido como o menor intervalo de acontecimento de um evento no sistema (por exemplo, adição ou remoção de um nó ou aresta). No entanto, diminuir a granularidade de observação implica em aumentar a complexidade na amostragem dos dados e na caracterização do sistema analisado. De forma análoga, quanto maior o valor de Δ , maior será o nível de agregação de informação incorporada ao grafo e menor a percepção da dinâmica associada.

4.2 ALGORITMOS DE DIFUSÃO BASEADOS EM MÉTRICAS DE CENTRALIDADE

Com o objetivo de acelerar o processo de difusão de informação em redes complexas que representam sistemas reais, considera-se o conhecimento de características das entidades que formam o sistema real. Abaixo está descrito o algoritmo proposto nesta dissertação,

tendo como objetivo a diminuição no tempo de difusão da informação. Este algoritmo faz uma análise das características topológicas dos nós, decidindo assim qual será a próxima entidade a receber a informação a ser difundida no sistema.

- i Seja o grafo $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$ que modela o sistema real.
- ii Seja o nó $v_t \in \mathcal{V}_t$ o dispositivo que possui a informação a ser difundida pela rede.
- iii Seja $\vartheta_t \subset \mathcal{V}_t$ o conjunto de dispositivos alcançáveis pelo nó v_t no instante de tempo t , isto é, o conjunto de vizinhos do nó v_t no instante t , dado pelos modelos CuT ou SoA.
- iv Em cada instante de tempo t , o nó v_t seleciona um elemento do conjunto ϑ_t com o maior valor da métrica de centralidade μ para o repasse da informação.

Nesta dissertação, as métricas μ utilizadas foram *Betweenness*, *Closeness* e Grau, dando origem às denominações *Betweenness Walk*, *Closeness Walk* e *Degree Walk* para os algoritmos propostos.

Para o conjunto de grafos \mathcal{G}_t , com $T_1 \leq t \leq T_N$, as métricas são calculadas em cada *snapshot*. Para o cálculo da métrica de *Betweenness*, *Closeness* ou Grau, supõe-se que o sistema possua uma entidade com visão global da topologia do grafo que modela o mesmo. Obviamente, para grafos com centenas de milhares de nós, o cálculo de *Betweenness* e *Closeness* pode ser custoso. Este trabalho pode ser estendido verificando a possibilidade do cálculo distribuído destas métricas. Por exemplo, para o cálculo de métricas de centralidade, os autores em Wehmuth e Ziviani (2011) apresentam um algoritmo que avalia de forma distribuída métricas de centralidade de maneira menos custosa.

4.2.1 CUSTOS DOS ALGORITMOS DE DIFUSÃO UTILIZANDO MÉTRICAS DE CENTRALIDADE

Nesta seção serão apresentados os custos dos algoritmos *Betweenness Walk*, *Closeness Walk* e *Degree Walk*, baseados no volume de tráfego gerado na rede. Dois tipos de informação são transmitidos pela rede:

- (i) Tráfego de dados, com volume de tráfego τ_d .
- (ii) Informações de controle, com volume de tráfego τ_c .

O tráfego de dados pode representar, por exemplo, um pedaço de um vídeo em uma aplicação par a par. Já as informações de controle são utilizadas pelos algoritmos para auxiliar no gerenciamento da difusão. No caso dos algoritmos propostos, as informações de controle serão representadas pelas medidas de centralidade de um nó. Claramente, $\tau_c \ll \tau_d$.

Em relação a estes métodos, a troca de dados entre os dispositivos acontece em duas etapas. Primeiramente, a vizinhança de um determinado nó envia sua medida de centralidade a ele. Será considerado que d_m é o tamanho médio da vizinhança dos nós. Desta forma, o volume de tráfego gerado nesta primeira etapa será dado por $\tau_c d_m$. Em segundo lugar, o nó vizinho com o maior valor da métrica de centralidade é escolhido e a mensagem é repassada a este. Sendo assim, o volume de tráfego gerado nesta segunda etapa será τ_d . Considere também que o tamanho médio de um caminho entre a origem e o destino na rede seja dado por π_a . Para estes algoritmos baseados em centralidade dos nós, o volume de tráfego será dado por $\mathcal{J}_c = \pi_a(\tau_d + \tau_c d_m)$. Para efeitos práticos em redes em larga escala, $\pi_a \ll |V|$. Sendo assim, é esperado que estes algoritmos gerem uma sobrecarga muito menor que os algoritmos epidêmicos. Além disso, a sobrecarga gerada pelos algoritmos orientados por centralidade, baseada no número de réplicas geradas na rede, está bem próxima a gerada pelos algoritmos baseados no passeio aleatório, conforme será mostrado no Capítulo 6.

4.3 UTILIZAÇÃO DAS MÉTRICAS EM UM RAIOS DE ABRANGÊNCIA

Nos algoritmos de difusão apresentados na seção anterior, ao se escolher qual vizinho receberá a informação, somente as métricas dos vizinhos diretos são consideradas. No entanto, é possível utilizar as métricas dos vizinhos pertencentes a um raio \mathcal{R} de abrangência (vizinhos de vizinhos), buscando uma melhor avaliação do próximo nó a receber a informação.

A Figura 4.4 mostra uma rede hipotética onde o vizinho com o valor do grau igual a 4 seria uma escolha pior para a difusão da informação do que o nó vizinho com valor de grau igual a 3.

Nesta proposta o algoritmo recebe um parâmetro \mathcal{R} que é definido como sendo o raio

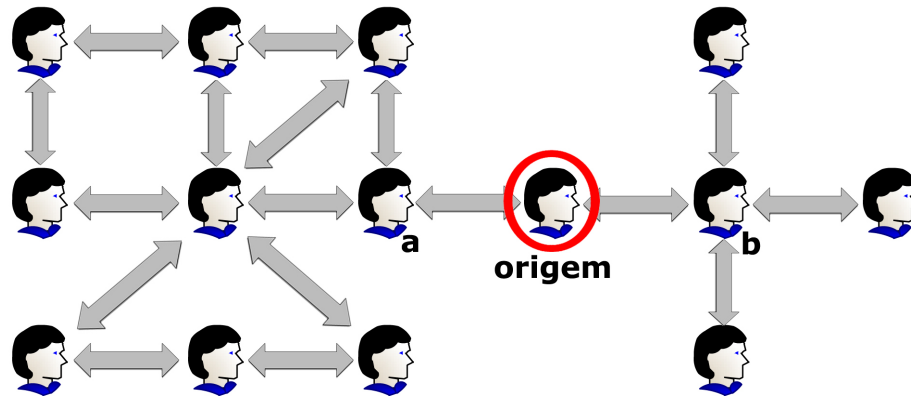


Figura 4.4: Centralidade dos vizinhos.

máximo da vizinhança em que as métricas são consideradas. A Figura 4.5 exemplifica a utilização de \mathcal{R} . Quando $\mathcal{R} = 1$, dada uma informação difundida pelo nó “Origem”, somente os valores das métricas dos vizinhos A e F seriam considerados. Porém, se o valor do parâmetro fosse $\mathcal{R} = 2$, as métricas dos vizinhos A, B, C, F, G, I também seriam consideradas. Da mesma forma, se $\mathcal{R} = 3$, são considerados todos os vizinhos de até três saltos da origem, o que faz com que todos os nós desta rede tenham suas métricas consideradas no cálculo.

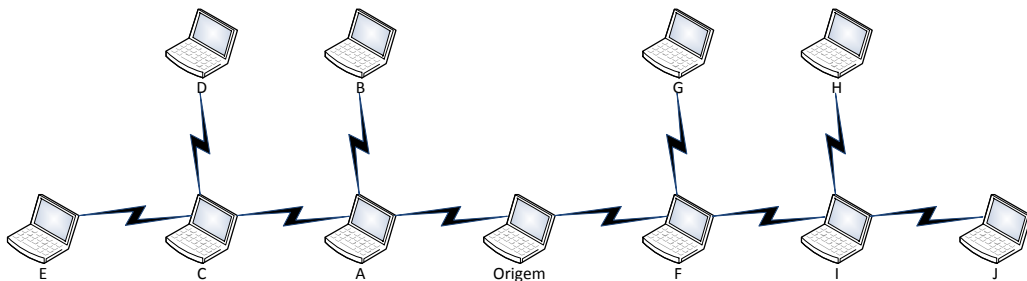


Figura 4.5: Utilização do valor da métrica dos vizinhos.

Esta proposta utiliza uma média ponderada de todos os vizinhos considerados na avaliação, encaminhando a informação para o vizinho que apresente uma maior média final. Esta média é ponderada por uma função de decaimento exponencial $f_d = \frac{1}{r}$, onde r corresponde ao raio do vizinho analisado, permitindo assim que vizinhos com um raio maior tenha um peso menor na média final. Em outras palavras, esta média faz com que os vizinhos mais próximos ao nó que difundirá a informação tenham um peso maior do que os vizinhos mais distantes.

4.4 MODELO DE PREVISÃO DE *LINKS*

Se for possível definir, com certa probabilidade, qual a formação topológica da rede no próximo instante de tempo, será possível utilizar essa informação na difusão da rápida de conteúdo pela rede. A proposta aqui descrita utiliza a previsão das arestas da rede no próximo instante de tempo para acelerar os processos de difusão da informação.

A previsão será feita baseando-se no histórico da rede. Em outras palavras, a medida com que os instantes de tempo vão passando, a probabilidade de existência de cada aresta é calculada levando em conta o histórico da formação topológica da rede.

Se os nós da rede pudessem prever que, em um futuro próximo, uma determinada ligação conveniente apareça, estes nós podem correr o risco de esperar até que a aresta seja formada, em vez de encaminhar a mensagem no instante atual. Agindo assim, o algoritmo aguardaria uma formação topológica mais favorável para a difusão da informação, ao invés de realizar o encaminhamento no instante de tempo atual.

Por exemplo, no cenário ilustrativo apresentado na Figura 4.6, o nó v deve enviar dados para o nó u . Durante o período inicial ($t = 0$), o nó v não tem alternativa a não ser transmitir dados para o nó a , esperando assim que os dados atinjam, eventualmente, o nó u em algum momento no futuro. No entanto, se o nó v pudesse prever que a aresta $v \rightarrow u$ existirá no futuro muito próximo (por exemplo, no instante $t = 1$, de acordo com a Figura 4.6(b)), o nó v pode então optar por esperar alguns instantes e entregar os dados diretamente para o nó u .

Se o nó v for incapaz de prever a existência da aresta $v \rightarrow u$ em um futuro próximo, a disseminação de dados para o nó u durará três unidades de tempo, considerando que cada encaminhamento de dados dure uma unidade de tempo. Em contrapartida, se o nó v prever a existência da aresta $v \rightarrow u$, a disseminação dos dados terá a duração de duas unidades de tempo. Porém, se o nó v falhar em sua previsão, a disseminação de informação será penalizada. Neste caso, o nó v esperaria quatro unidades de tempo até que os dados cheguem ao nó u .

Com base neste exemplo simples, é evidente que um mecanismo de previsão das arestas existentes nos grafos subsequentes deve ser simples o suficiente para ser eficaz, além também de ser preciso o suficiente para que se tenha a taxa de erro mais baixa possível. Nesta dissertação é proposto um mecanismo de previsão, tendo como base um Modelo Markoviano de dois estados, onde uma determinada ligação existe no estado E e não existe

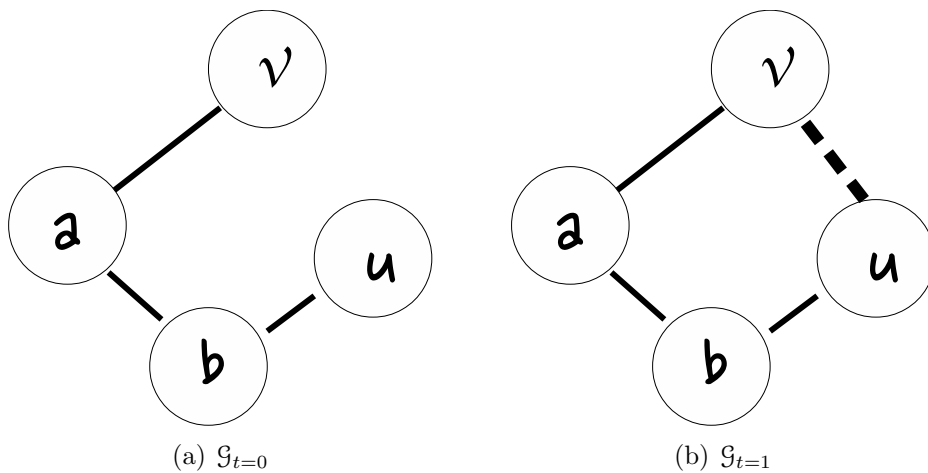


Figura 4.6: Dois *snapshots* consecutivos ilustrando o aparecimento de uma ligação.

no estado $!E$. A Figura 4.7 ilustra o mecanismo de previsão de arestas. De acordo com este modelo, a conexão $v \rightarrow u$ pode existir ou não, durante um determinado *snapshot*.

Se, no instante de tempo t a aresta existir, ele não existirá no momento $t + 1$ com probabilidade igual a λ e continuará existindo no instante $t + 1$ com probabilidade $1 - \lambda$. Da mesma forma, se a ligação não existir no instante t , ela existirá no instante $t + 1$ com probabilidade igual a ρ e continuará inexistente com probabilidade $1 - \rho$. Sendo assim, se o nó v for capaz de prever a existência de uma aresta, este poderá considerar o nó u como um dos candidatos a receber os dados no próximo instante de tempo.

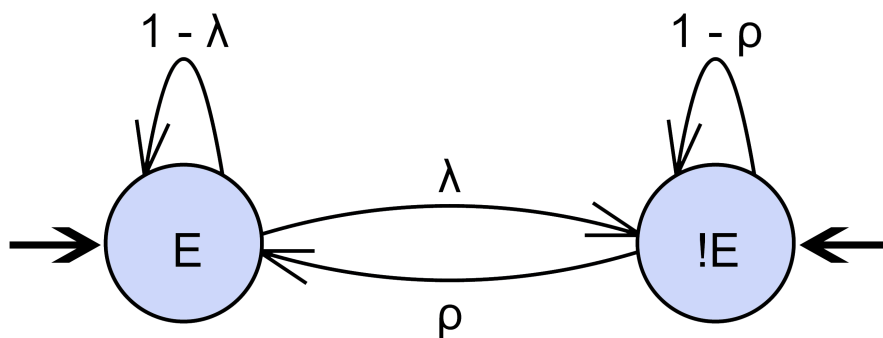


Figura 4.7: Modelo para o estado de *link* dinâmico.

Para se ter uma visão temporal da evolução do estado de cada aresta, é definida a

função $E_{v \rightarrow u}^t$ como sendo a existência da aresta $v \rightarrow u$ no tempo t , da forma como segue.

$$E_{v \rightarrow u}^t = \begin{cases} 1, & v \rightarrow u \in \mathcal{E}_t. \\ 0, & \text{caso contrário.} \end{cases} \quad (4.1)$$

Seja $E_{v \rightarrow u}^{t+1}$ a representação da existência da aresta $v \rightarrow u$ no próximo *snapshot*, isto é, em \mathcal{G}_{t+1} . Desta forma é possível calcular $E_{v \rightarrow u}^{t+1}$, tendo como base o modelo *Weighted Moving Average*, apresentado em Roberts (1959), utilizando para isso os últimos b valores amostrados da aresta $v \rightarrow u$. Calcula-se uma média ponderada, onde os pesos maiores são atribuídos aos valores de $E_{v \rightarrow u}^t$ que possuem os instantes t mais recentes, amostrados em uma janela de tamanho b . Cada valor passado do estado de *link* ($E_{v \rightarrow u}^t, E_{v \rightarrow u}^{t-1}, \dots, E_{v \rightarrow u}^{t-b}$) é ponderado por um fator α , fator este que tem um decaimento exponencial. Desta forma, a Equação 4.2 apresenta o cálculo da probabilidade prevista de $E_{v \rightarrow u}^{t+1}$:

$$P(E_{v \rightarrow u}^{t+1}) = \frac{E_{v \rightarrow u}^t + E_{v \rightarrow u}^{t-1} * \alpha + E_{v \rightarrow u}^{t-2} * \alpha^2 + \dots + E_{v \rightarrow u}^{t-b} * \alpha^b}{1 + \alpha + \alpha^2 + \dots + \alpha^b}, \quad (4.2)$$

O valor de α ($0 < \alpha < 1$) é um fator que pondera a média, decaindo exponencialmente. Neste sentido, os valores amostrados mais distantes do instante atual de observação rapidamente têm sua importância reduzida na Equação 4.2. A partir dessa sequência de valores b amostrados, é possível também parametrizar λ e ρ no modelo da Figura 4.7. Cabe ainda dizer que as métricas utilizadas na difusão da informação são agora calculadas no grafo previsto. Isto faz com que a informação chegue ao nó que possui o maior valor da métrica no instante seguinte, isto é, chegue ao nó com maiores chances de difundir a informação no próximo salto.

Desta forma, a previsão de arestas será utilizada na busca da difusão rápida da informação pela rede, proporcionando assim um ganho alcançado pela escolha do melhor indivíduo para o encaminhamento da informação na rede.

4.5 MEDIDA DE INTERESSE

Em todos os resultados relativos aos algoritmos de difusão de informação considera-se a medida de interesse *percentual de nós alcançados em t*, apresentada na equação 4.3.

$$\eta(t) = \frac{\sum_{\forall v \in \mathcal{V}_t} \tau_t(v)}{|\mathcal{V}_t|^2}. \quad (4.3)$$

Nesta equação, $\tau_t(v)$ representa o total de nós alcançados a partir do nó v até o *snapshot* t , $T_1 \leq t \leq T_N$. Esta métrica calcula a média dos nós alcançados a cada grafo $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$, considerando cada um dos nós em \mathcal{V}_t como nó inicial de difusão. Como é buscado o percentual dos nós alcançados, a métrica do valor médio é normalizada. Vale ressaltar que para não introduzir *bias* na escolha do nó inicial, levando a um caminho aleatório que favoreça a um determinado algoritmo de difusão, opta-se pela média calculada considerando todos os nós como sendo nós iniciais da difusão. É importante dizer também que um nó é considerado como alcançado no primeiro instante de tempo que este evento ocorre.

5 CONJUNTOS DE DADOS UTILIZADOS

Neste capítulo será apresentado o conjunto de redes utilizado para a análise dos algoritmos de difusão estudados. As redes utilizadas representam diferentes sistemas reais, bem como os principais modelos teóricos encontrados na literatura.

Os dados que representam sistemas reais são descritos a seguir. O primeiro reúne informações sobre a mobilidade de usuários entre diferentes pontos de acesso de uma rede *wireless* de um *campus* universitário (KOTZ et al., 2009). O segundo reúne informações sobre a mobilidade de usuários carregando pequenos dispositivos (*iMotes*) durante o *Info-com 2006* (SCOTT et al., 2009). O terceiro representa as trocas de informações entre os diversos usuários de uma aplicação *P2P* de difusão de vídeo ao vivo (FERREIRA et al., 2013). Os dois primeiros sistemas se enquadram em sistemas tolerantes a atraso: o importante é a entrega, por exemplo, de uma atualização de software ao maior número possível de usuários. No terceiro sistema, o mais importante é a entrega rápida da informação, mesmo que alguns usuários não a recebam.

Já em relação às redes sintéticas, foram utilizados os modelos de redes complexas apresentados no Capítulo 2 para a geração de cada um dos *snapshots*. Foi utilizado também um gerador de redes sintéticas baseado no modelo *Edge-Markovian*, buscado assim acrescentar dependência estrutural entre os *snapshots*, característica esta inexistente nas redes sintéticas geradas, conforme apresentado na Seção 5.4.

5.1 DARTMOUTH

O *log* Dartmouth foi apresentado em Kotz e Essien (2002) e inclui informações do protocolo SNMP (CASE et al., 1989; FEIT, 1993) de pontos de acesso da rede *wireless* no campus Dartmouth College, no período entre abril de 2001 e junho de 2004. 13.888 dispositivos utilizaram a rede nos 4 anos de coleta. Este *log* tem sido amplamente utilizado em diversos trabalhos na literatura que tratam da mobilidade de usuários de redes *wireless* e da troca de informação entre eles, como por exemplo, Barbera et al. (2011).

Para representar os contatos entre os usuários a partir dos dados disponibilizados, é seguida a metodologia apresentada em Chaintreau et al. (2007): Assume-se que dispositivos associados a um mesmo ponto de acesso em um dado intervalo de tempo estão

conectados. Em termos da representação matemática, estes dispositivos são vizinhos no grafo, ou seja, existem arestas que os conectam. Todos os nós associados a um mesmo ponto de acesso em um dado intervalo de tempo formam um grafo totalmente conectado. Este grafo é gerado de uma forma não direcionada, indicando que, se dois nós estão conectados a um mesmo ponto de acesso, existe então caminho para troca de informação em ambas as direções.

Nos resultados apresentados nesta dissertação, é selecionado o período de um ano (janeiro a dezembro de 2002), com a presença de 5.316 dispositivos. O ano de 2002 foi escolhido dado que este foi o primeiro ano completo de coleta de informações entre os dispositivos. Para a construção do grafo \mathcal{G}_t , a escolha para o valor Δ foi de 3.600 segundos (em linha com resultados mostrados em Balachandran et al. (2002)). Assim, um total de 8.769 *snapshots* diferentes foi analisado.

5.2 INFOCOM

O *log Infocom* foi apresentado por Scott et al. (2009) e inclui informações de visualizações *Bluetooth* (BLUETOOTH, 2001) transportando pequenos dispositivos (iMotes) por quatro dias, durante a conferência *Infocom*, em 2006, na cidade de Barcelona, Espanha. A coleta foi realizada entre os dias 24 de Abril e 27 de Abril de 2006.

O *log Infocom* é um exemplo clássico de redes DTNs (FALL; FARRELL, 2008), onde os membros integrantes armazenam a informação recebida e a repassa somente em instantes subsequentes, quando novas parcerias são formadas devido às movimentações dos participantes durante o evento. O *log* conta com dezessete iMotes estáticos, de longo alcance, cobrindo toda a área da conferência. Outros três iMotes, também de longo alcance, foram implantados no elevador do local onde foi realizada a conferência. Os demais dispositivos, somando setenta e oito iMotes, foram distribuídos a membros participantes do Infocom, totalizando noventa e oito dispositivos com informações coletadas no experimento.

Para a modelagem deste sistema, foram consideradas que cada visualização entre qualquer par de dispositivos gera uma aresta entre estes membros. Para a construção do grafo \mathcal{G}_t , a escolha dos valores Δ foi de 60 segundos e 900 segundos, gerando um total de 5.624 e 375 *snapshots* diferentes a serem analisados, em linha com as modelagens apresentadas em Pietiläinen e Diot (2012) e Mtibaa et al. (2010).

5.3 SOPCAST

O SopCast (SOPCAST, 2013) é uma das aplicações mais populares para difusão de vídeo em tempo real baseada em redes *P2P*. Cada canal transmitido pelo SopCast possui uma rede sobreposta própria, com a transmissão de um vídeo particular, isto é, uma rede lógica é formada para transmitir conteúdos específicos. Esta rede sobreposta para transmissão do vídeo está relacionada com a aplicação através de um canal no qual um usuário (cliente) está conectado. Clientes SopCast que utilizam a aplicação estão conectados a um canal em particular, pertencendo a rede sobreposta na qual o vídeo está sendo transmitido.

O *log* modelado neste trabalho foi apresentado em Ferreira et al. (2013) e representa uma coleta de 1h realizada no dia 14 de outubro de 2011. Para a construção do grafo \mathcal{G}_t , considera-se o valor de $\Delta = 1$ segundo, capturando uma quantidade significativa de troca de informação entre vizinhos. Implicitamente está sendo capturando neste *snapshot* a dinâmica de descoberta de novas parcerias que estabelecem troca de informação. São considerados 3.600 *snapshots* de 1 segundo, com o total de 334 nós no grafo que modela o sistema, em linha com a modelagem apresentada em Tang et al. (2009) e Gonçalves et al. (2012). Uma aresta é estabelecida entre dois nós da rede sobreposta se existir, pelo menos, uma troca de pacote maior que 200 *bytes*, considerado pacote de vídeo (descartando pacotes de controle) (TANG et al., 2009). O grafo que representa matematicamente o sistema é um grafo direcionado, indicando assim o sentido da troca de pacotes de vídeo entre os participantes.

5.4 REDES SINTÉTICAS

Para complementar a análise dos algoritmos de difusão propostos, foram criadas redes seguindo os modelos Erdős e Rényi, Watts e Strogatz (conhecido como modelo *Small World*) e também Barabási - Albert, ambos apresentados no Capítulo 2. Para a geração destas redes, foi utilizada a ferramenta *NetworkX - High-Productivity Software for Complex Networks*, apresentada em Hagberg et al. (2008).

A Tabela 5.1 apresenta a parametrização utilizada na geração dos grafos para cada um dos modelos encontrados na literatura. Para cada um destes parâmetros, foram gerados 100 grafos sintéticos. No modelo Erdős e Rényi, o parâmetro p indica a probabilidade

de existência de uma aresta entre cada um dos pares de nós da rede. Já o modelo *Small World* é iniciado com um grafo em anel, onde cada nó é simetricamente ligado aos k vizinhos mais próximos. A partir desta formação, para cada nó, cada ligação deste nó com seu vizinho no sentido horário é retirada com probabilidade p e reconectada a um nó aleatório. O modelo Barabási - Albert é iniciado com m nodos isolados. A cada instante de tempo um novo nó j é adicionado no grafo com $m_j \leq m$ conexões.

Parâmetros	Erdős e Rényi	<i>Small World</i>	Barabási - Albert
p_1	0.01	0.01	-
p_2	0.05	0.05	-
p_3	0.1	0.1	-
k_1	-	10	-
k_2	-	50	-
k_3	-	100	-
m_1	-	-	10
m_2	-	-	50
m_3	-	-	100

Tabela 5.1: Parâmetros utilizados na geração das redes sintéticas.

Todas as redes sintéticas possuem o total de 1000 nós. Para a criação do grafo temporal \mathcal{G}_t , considera-se o valor de $\Delta = 1$ unidade de tempo, o que corresponde a um grafo sintético gerado pela ferramenta.

5.5 MODELO *EDGE-MARKOVIAN*

Em sistemas reais, ao observar *snapshots* em instantes subsequentes, é possível notar certa dependência do instante atual com os instantes anteriores (NEWMAN, 2003). Essa dependência pode ser maior ou menor, dependendo do sistema observado, mas está sempre presente. Uma forma de simular este comportamento é utilizar modelos que geram o instante atual a partir de características do sistema, presentes no instante anterior.

A partir de um grafo inicial, é possível definir grafos subsequentes seguindo o modelo *Edge-Markovian* presente na literatura em Eichel e Delp (1984); Coolen et al. (2009). O grafo atual é gerado a partir do Modelo Markoviano de definição de arestas, onde a existência de cada aresta é definida por uma Cadeia de Markov, tendo como base as arestas do grafo anterior.

A Cadeia de Markov utilizada na geração é análoga é apresentada na Seção 4.4 e possui

os mesmos parâmetros, λ e ρ , conforme apresentado na Figura 5.1. Porém, neste caso a cadeia é utilizada na geração sintética da rede, diferente do apresentado na Seção 4.4, onde a cadeia é utilizada na previsão do estado das arestas.

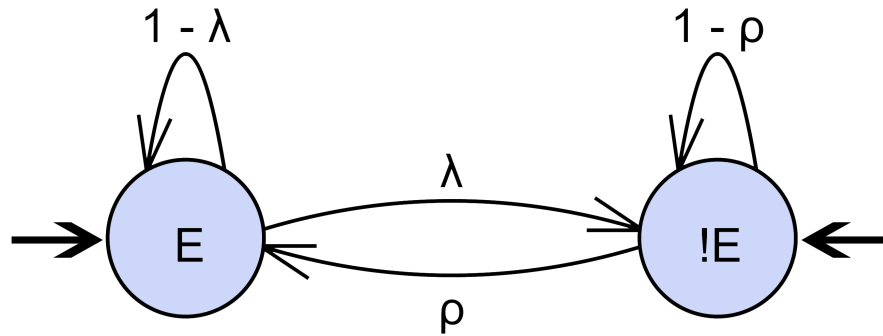


Figura 5.1: Cadeia de Markov utilizada no gerador.

	Baixa Dinâmica	Média Dinâmica	Alta Dinâmica
Parâmetro λ	0.1	0.5	0.9
Parâmetro ρ	0.1	0.5	0.9

Tabela 5.2: Parâmetros utilizados pelo Gerador Markoviano.

Neste trabalho foram geradas redes com este Gerador Markoviano, utilizando para isso grafos iniciais gerados com cada um dos modelos teóricos Erdős e Rényi, *Small World* e Barabási - Albert. Para cada um dos modelos, foram criadas três redes distintas, ambas com 1000 nós, utilizado para isso três níveis de dinâmica, dada pela parametrização dos valores de λ e ρ , conforme apresentada na Tabela 5.2. Para cada rede foram gerados 50 grafos (*snapshots*) com valor de $\Delta = 1$ unidade de tempo, o que corresponde a 1 grafo sintético gerado, assim como na Seção 5.4. Desta forma, foram criadas redes sintéticas com um comportamento aproximado ao comportamento de sistemas reais, mantendo dependência do instante atual com o histórico da rede.

6 RESULTADOS

O foco dos resultados alcançados nesta dissertação é avaliar os ganhos alcançados ao utilizar as métricas de centralidade para acelerar os processos de difusão em redes dinâmicas. Será avaliado também impacto ao considerar a dinâmica inerente aos sistemas reais na modelagem de tais sistemas. Este impacto é mensurado na avaliação dos algoritmos de difusão em redes complexas dinâmicas.

6.1 ANÁLISE DOS MODELOS SOA E CUT

Como primeiro conjunto de resultados, é apresentado o impacto da dinâmica de cada sistema na difusão de informação para a implementação do algoritmo Epidêmico nos modelos SoA e CuT. A dinâmica dos sistemas será modelada como inserção, permanência ou retirada de arestas em cada um dos *snapshots* do grafo \mathcal{G}_t que representa o sistema.

A dinâmica pode ser observada de diferentes formas. Por exemplo, entrada e saída de pares em um canal de compartilhamento de vídeo, mudança da localização geográfica de dispositivos móveis, acarretando em alteração da estação base em que tal dispositivo está associado, movimentação de participantes em algum evento, dentre outras.

Para o *log Dartmouth* a dinâmica será definida como mudança de localização de um dispositivo entre os diferentes pontos de acessos que formam a rede *wireless*. Já para o *log* da aplicação SopCast, a dinâmica do sistema está relacionada com a criação de novas parcerias para troca de porções de vídeo, geradas pelo protocolo da aplicação.

A Figura 6.1 apresenta o resultado da difusão da informação entre as entidades destes dois sistemas reais. O tempo necessário para atingir um percentual significativo de nós ($> 20\%$) do cenário *Dartmouth* é três ordens de grandeza maior que a duração de um *snapshot* ($\Delta = 1h$), para ambos os modelos CuT e SoA. Em contrapartida, para o caso da aplicação SopCast, 80% dos nós são alcançados em um tempo igual a 10 vezes a duração de um *snapshot* ($\Delta = 1s$). Este resultado está relacionado com características de cada um dos sistemas: enquanto a aplicação SopCast é “gulosa” na descoberta de novas parcerias, usuários de redes *wireless* se movimentam entre áreas de cobertura seguindo um comportamento cotidiano (seus usuários não estão ávidos a se conectar ou “conhecer” todas as entidades do sistema, ou seja todos os pontos de acesso). É possível notar também

um resultado semelhante entre os modelos SoA e CuT para o *log Dartmouth*. Este fenômeno ocorre dado que todos os nós dentro de uma área de cobertura de um ponto de acesso estão conectados entre si, formando um grafo completamente conectado, seguindo a modelagem apresentada em Chaintreau et al. (2007). Desta forma, os modelos SoA e CuT se comportam de forma análoga, atingindo todos os nós dentro de uma componente ligada. Tal característica não é encontrada no SopCast, pois trata-se de uma aplicação onde cada usuário se conecta ao dispositivo que possui a porção do vídeo necessária em cada instante.

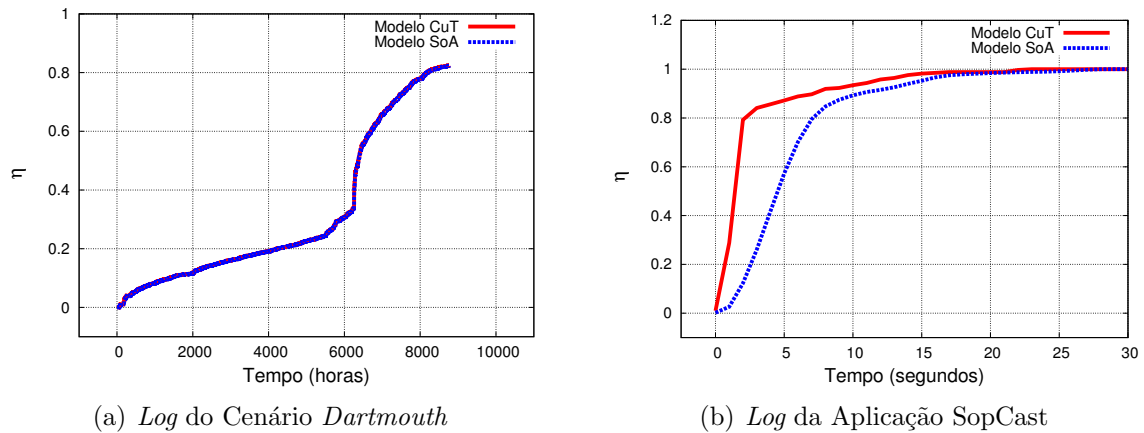


Figura 6.1: Dinâmica considerando mobilidade: Desempenho dos modelos SoA e CuT (**Epidêmico**).

6.2 IMPACTO DA DINÂMICA NO PROCESSO DE DIFUSÃO

Neste conjunto de resultados é analisada a forma com que a avaliação dinâmica impacta nos resultados dos algoritmos de difusão, utilizando o conjunto de dados do SopCast. A dinâmica é uma característica presente nos sistemas reais e nem sempre ela é considerada nos algoritmos de difusão, conforme apresentado no Capítulo 3. O tempo de cobertura da rede para o grafo estático G é muito rápido, isto porque todas as arestas que são formadas durante todo o período de observação já estão disponíveis desde o primeiro passo da difusão. Usando o algoritmo Epidêmico juntamente com o modelo SoA, 100% dos nós são atingidos no segundo instante de tempo, conforme mostrado na Figura 6.2. O algoritmo *Random Walk* necessita de 8 segundos para difundir a informação no grafo estático G . Porém, conforme a Figura 6.3(a) discute adiante, é possível perceber que o tempo para difusão da informação é três vezes maior quando a dinâmica do sistema é

considerada na modelagem. Para os algoritmos propostos, o tempo de cobertura fica 20 vezes menor quando a dinâmica do sistema é desconsiderada. Na maioria dos conjuntos de sistemas reais, é importante estimar o tempo de difusão da forma mais exata possível (YOSHIDA, 1992).

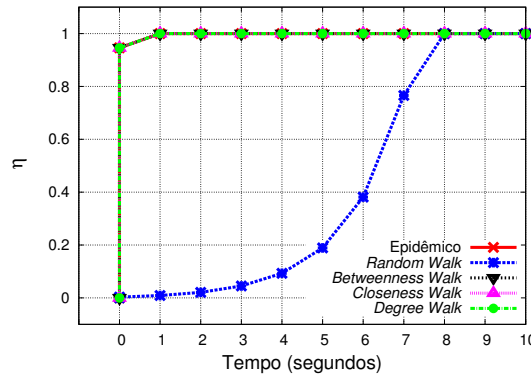


Figura 6.2: Algoritmos de difusão no agregado - Modelo SoA.

6.3 DIFUSÃO ORIENTADA POR CENTRALIDADE

É importante ressaltar que não somente a dinâmica influencia o tempo de difusão. A Figura 6.3 mostra que o uso das métricas de centralidade para as decisões de encaminhamento também melhoram o processo de difusão. Tanto no modelo SoA quanto no CuT, ao utilizar métricas de centralidade ocorre uma diminuição no tempo de difusão para a mesma porção de nós alcançados, quando comparados ao resultado de um simples passeio aleatório. Em outras palavras, a escolha dos nós mais centrais acelera o processo de difusão, considerando o tempo total para difundir informação sobre uma porção de nós. Além disso, conforme esperado, os melhores resultados quanto ao tempo de cobertura são alcançados pelo algoritmo Epidêmico, mas tal algoritmo possui um custo elevado.

Para a obtenção do valor das métricas foi utilizada a ferramenta *NetworkX - High-Productivity Software for Complex Networks*, apresentada em Hagberg et al. (2008), onde é considerada a Componente Gigante¹ do grafo, evitando assim que nós desconexos gerem valores infinitos ou nulos.

Serão considerados os resultados referentes ao modelo *Store-or-Advance* apresentados na Figura 6.3(a), calculados nos *logs* da aplicação SopCast. Conforme esperado, o melhor desempenho é alcançado com o algoritmo Epidêmico, que abrange todos os vizinhos

¹Componente Gigante é definida como a componente conectada que contém a maior parte dos nós na rede.

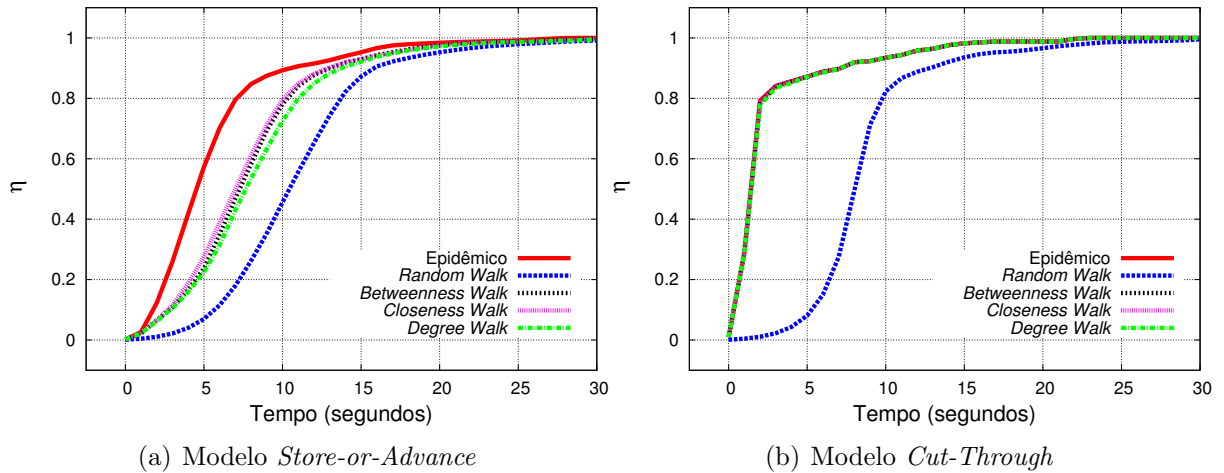


Figura 6.3: Difusão orientada por centralidade - SopCast.

a cada iteração do algoritmo. A consideração mais interessante refere-se à diferença do tempo de difusão entre os algoritmos *Random Walk*, *Betweenness Walk*, *Closeness Walk* e *Degree Walk*: ao se incluir as características de centralidade dos nós no processo de difusão, diminui-se em aproximadamente 20% o tempo para difundir a informação a 50% dos nós da rede, para todos os algoritmos dirigidos por centralidade, quando comparado ao algoritmo *Random Walk*. A pequena diferença entre os algoritmos pode ser explicada pela característica desta rede em particular: o diâmetro² desta rede é pequeno, aproximadamente igual a 4. Assim pode-se dizer que a maioria dos nós está próximo aos demais membros da rede. Sendo assim, os nós pertencem a maioria dos caminhos mínimos, resultando em um pequeno ganho do algoritmo *Closeness Walk*. Outra característica que influencia os resultados alcançados é o fato da formação da rede do SopCast se aproximar de uma topologia aleatória.

A Figura 6.3(b) apresenta os resultados referentes ao modelo *Cut-Through*. Ao se considerar as métricas *Betweenness*, *Closeness* e Grau no processo de difusão, o desempenho se aproxima ao desempenho do algoritmo Epidêmico. Alguns pontos explicam estes resultados. O modelo CuT permite a difusão da informação a todos os nós que pertencem a componente ligada, incluindo aquele que atinge o maior número de nós em um menor número de saltos. A escolha entre o nó mais próximo de todos os demais nós da rede e o nó que está na maior fração de caminhos mínimos entre quaisquer dois nós é equivalente, dada a característica do pequeno diâmetro da rede.

Quanto mais simples o cálculo da métrica considerada na difusão, menor a sobrecarga

²diâmetro é definido como o maior entre todos os caminhos mínimos

do sistema para se obter tais valores. Sendo assim, em sistemas reais pode ser necessário utilizar métricas que possuem menor complexidade computacional, devido ao número elevado de nós presentes. Ao utilizar a informação do Grau para a difusão, ocorre também uma diminuição no tempo de difusão da informação. Pela Figura 6.3 é possível notar que o algoritmo *Degree Walk* se comporta de forma semelhante aos algoritmos mais custosos como *Betweenness Walk* e *Closeness Walk*. Como o valor do Grau de um nó é obtido de forma mais simples que as demais métricas, têm-se um reforço da ideia de que é vantajoso adotar a medida de centralidade que possui o cálculo com a menor complexidade computacional (VALENTE et al., 2008).

Assim como a utilização de métricas de centralidade aceleram o processo de difusão de informação, é importante também analisar como o conhecimento do sistema contribui para melhorar a própria difusão. De acordo com a modelagem definida neste trabalho, o conhecimento do sistema é proporcional ao tamanho *snapshot* Δ , usado para modelar o conjunto de grafos \mathcal{G}_t que representam o sistema de forma dinâmica.

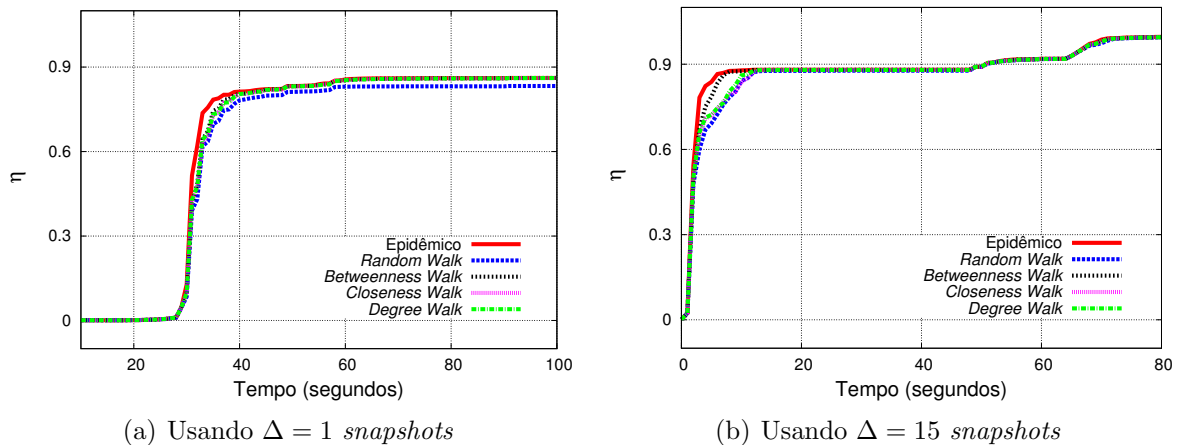


Figura 6.4: Impacto do conhecimento do sistema (Δ) na proporção dos nós alcançados.

A partir deste ponto, somente o modelo SoA é considerado devido ao fato dos resultados dos algoritmos de difusão no modelo CuT serem equivalentes ao Epidêmico. Isto ocorre uma vez que, como o grafo \mathcal{G}_t das redes consideradas é denso, o modelo CuT já permite uma difusão equivalente ao Epidêmico, independente da métrica considerada, conforme visto no gráfico 6.3(b).

Para o conjunto de dados Infocom, as Figuras 6.4(a) e 6.4(b) mostram que modelar o sistema com $\Delta = 15$ o tempo de cobertura fica menor quando comparado com o mesmo sistema, porém modelado com $\Delta = 1$. Isto ocorre porque existem mais arestas disponíveis

em cada *snapshot*. Ocorre ainda um aumento no número total de usuários que recebem a mensagem de 85% para 100%. Para o cenário modelado com $\Delta = 15$, o *Betweenness Walk* se aproxima melhor do algoritmo Epidêmico, espalhando mais rapidamente a informação, quando comparado com os demais algoritmos baseados em centralidade.

Neste sistema em particular, o maior conhecimento do sistema em cada instante (ou seja, maior Δ) ajuda na escolha dos participantes-chave com uma maior centralidade. Estes participantes atuam como pontes que ligam diferentes grupos de usuários na rede e pode ser interpretado como os participantes mais populares, conhecendo assim um grande número de pessoas. Além disso, eles tendem a possuir alta mobilidade, atingindo os 100% dos participantes registrados na conferência.

6.4 REDES SINTÉTICAS

A seguir são apresentados os resultados dos algoritmos de difusão, aplicados nas redes sintéticas apresentadas na Seção 5.4. Certas características de redes reais como coeficiente de clusterização, distribuição do grau dos nós e o diâmetro da rede são bem conhecidas. Diante disso, ao utilizar os modelos teóricos na geração das redes sintéticas, é possível simular a formação topológica de redes reais. Para cada conjunto \mathcal{G}_t gerado, cada grafo é uma instância gerada de forma independente pelo modelo, não existindo assim qualquer histórico ou dependência entre o grafo g_t e g_{t+1} , onde $\{g_t; g_{t+1}\} \subset \mathcal{G}_t$. Os parâmetros utilizados na geração das redes sintéticas estão apresentados na Tabela 5.1.

Serão discutidos os resultados com base nas características presentes nos modelos teóricos utilizados na a geração das redes. Conforme esperado, para o modelo Erdős e Rényi (Figura 6.5), aplicar os algoritmos baseados em métricas de centralidade não aceleram o processo de difusão quando comparados com o algoritmo *Random Walk*. A razão para este comportamento se baseia na homogeneidade dos nós, principalmente com base na distribuição de *Poisson* do grau dos nós.

No modelo Barabási - Albert, como é possível ver na Figura 6.6, a difusão orientada por centralidade obteve ganhos em relação à difusão *Random Walk*. Além disso, as métricas de centralidade obtiveram resultados equivalentes. Sendo assim, a escolha de qualquer métrica para orientar difusão levará aos mesmos resultados. O *Coefficiente de Correlação de Spearman*³ que mensura a dependência estatística entre duas variáveis, apresentado na

³O *Coefficiente de Correlação de Spearman* é uma medida de correlação não paramétrica que avalia a

Tabela 6.1, colabora com esta afirmação. A tabela apresenta os valores máximos e mínimos da correlação dos pares de métricas, quando esta é calculada em todas as entidades da rede. Todas as métricas de centralidade estão fortemente correlacionadas.

Modelo	<i>Bet-Clos</i>		<i>Bet-Deg</i>		<i>Clos-Deg</i>	
	Min.	Max.	Min.	Max.	Min.	Max.
Barabási-Albert	0.8849	0.9329	0.9693	0.9794	0.8812	0.9537

Tabela 6.1: Coeficiente de Correlação de Spearman.

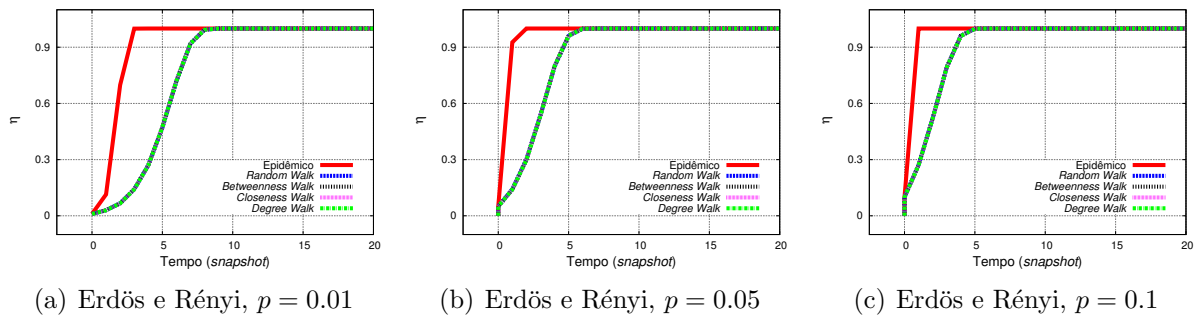


Figura 6.5: Difusão nos grafos sintéticos gerados com o modelo Erdős e Rényi.

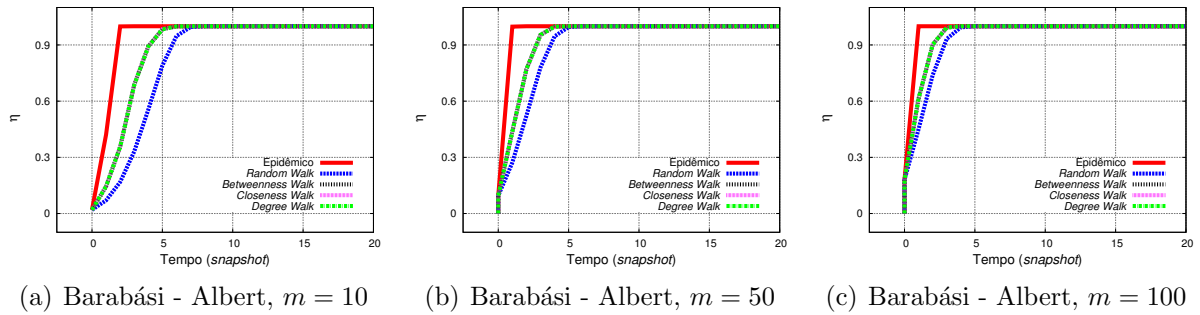
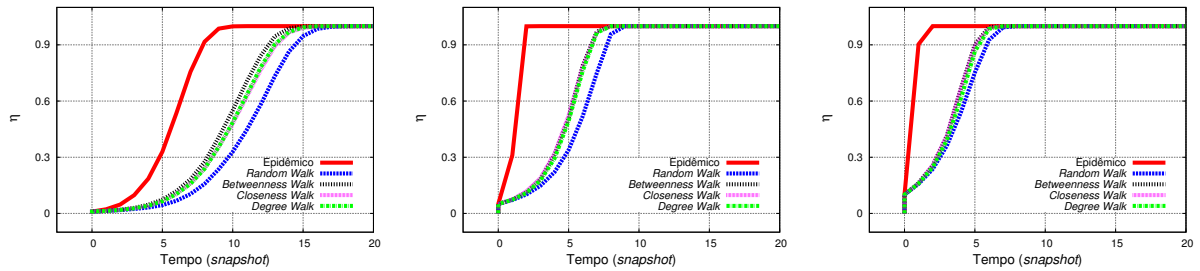


Figura 6.6: Difusão nos grafos sintéticos gerados com o modelo Barabási - Albert.

Para o modelo *Small World*, os algoritmos de difusão se comportam de maneira diferente. O processo de difusão é mais sensível às métricas de centralidade. Como é possível notar na Figura 6.7, para o valor $p = 0.01, k = 10$, os algoritmos de difusão baseados nas métricas de Grau e *Closeness* apresentam o mesmo padrão de difusão. Entretanto, ao utilizar *Betweenness* para orientar a difusão é possível perceber um resultado melhor.

dependência estatística de duas variáveis. Além do mais, ela avalia a correlação através de uma função monótona arbitrária que pode ser a descrição da relação entre duas variáveis, sem fazer nenhuma suposição sobre a distribuição de frequências delas. Neste trabalho é utilizada esta correlação por ela ser pouco afetada por valores discrepantes.

Já para os valores $p = 0.05, k = 50$ e $p = 0.1, k = 100$ a difusão por métrica manteve o mesmo padrão, não importando a métrica escolhida.



(a) *Small World*, $p = 0.01, k = 10$ (b) *Small World*, $p = 0.05, k = 50$ (c) *Small World*, $p = 0.1, k = 100$

Figura 6.7: Difusão nos grafos sintéticos gerados com o modelo *Small World*.

Estes resultados demonstram que processos de difusão são acelerados quando características de centralidade das redes complexas são utilizadas, com exceção das redes com lei de formação aleatória.

6.5 MODELO *EDGE-MARKOVIAN*

Nesta seção serão apresentados os resultados dos algoritmos de difusão *Betweenness Walk*, *Closeness Walk* e *Degree Walk* quando aplicados em redes sintéticas geradas a partir do modelo *Edge-Markovian* apresentado na Seção 5.5. Tal Modelo Markoviano foi utilizado para que seja incluído um histórico entre *snapshots* consecutivos. Em outras palavras, a intuição por trás da aplicação deste modelo é que exista uma dependência entre os grafos de instantes subsequentes, dependência que não existe nas redes sintéticas geradas na seção anterior.

A Tabela 5.2 apresenta os valores utilizados nos parâmetros λ e ρ na Cadeia de Markov do modelo. Estes parâmetros indicam a permanência ou alteração nos estados da cadeia, indicando a existência ou não de uma aresta no grafo gerado. Para cada um dos modelos de redes complexas apresentados no Capítulo 2 (Erdős e Rényi, *Small World* e Barabási - Albert) foram gerados três conjuntos de 50 grafos sintéticos em cada um deles. Cada conjunto gerado equivale a uma combinação de parâmetros da Tabela 5.2, correspondendo assim às graduações alta, média e baixa dinâmica.

As Figuras 6.8, 6.9 e 6.10 apresentam os resultados dos algoritmos de difusão, quando aplicados nos grafos gerados com os parâmetros $\lambda = \rho = 0.1$, $\lambda = \rho = 0.5$ e $\lambda =$

$\rho = 0.9$ respectivamente. Serão apresentados somente os resultados do modelo *Small World*, visto que para os modelos Erdős e Rényi e Barabási - Albert, o comportamento é similar. É possível notar que os resultados do processo de difusão nos *logs* gerados pelo Modelo Markoviano tiveram o comportamento próximo aos resultados da difusão nas redes sintéticas geradas pelo modelo Erdős e Rényi. Desta forma, os algoritmos de difusão baseados em centralidade obtiveram um comportamento equivalente ao algoritmo *Random Walk*.

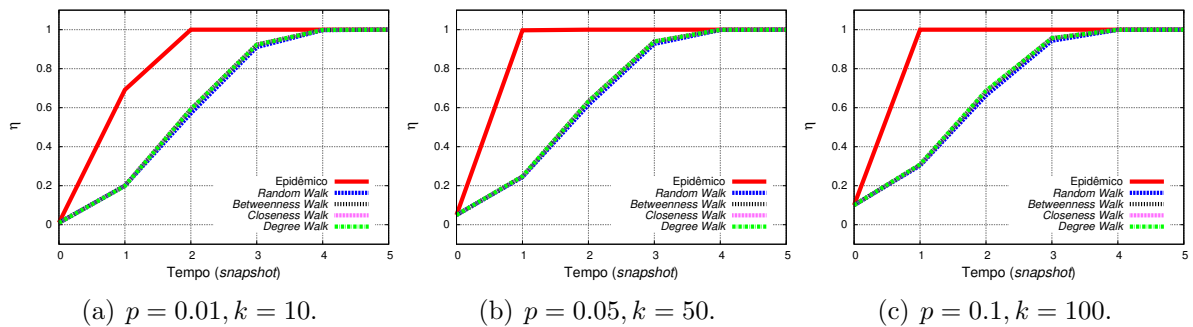


Figura 6.8: *Small World* - Baixa Dinâmica.

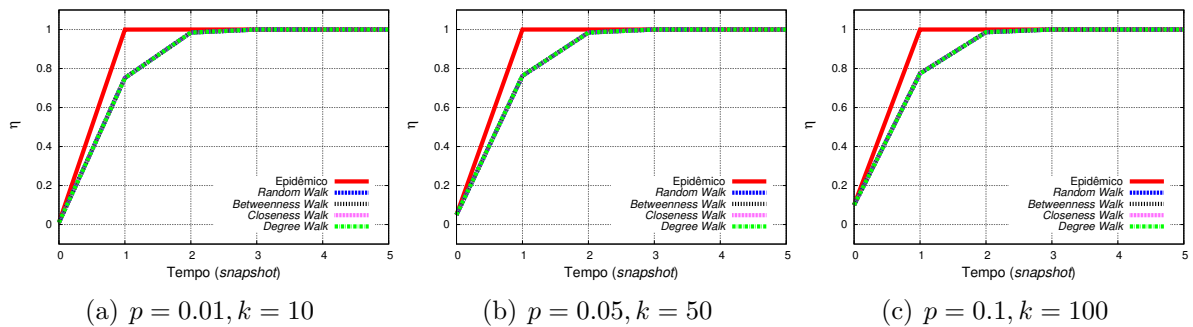


Figura 6.9: *Small World* - Média Dinâmica.

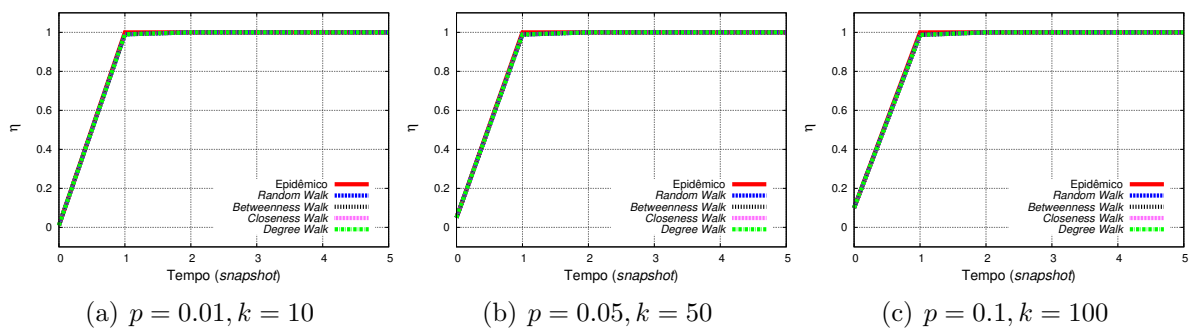


Figura 6.10: *Small World* - Alta Dinâmica.

Ao utilizar o Gerador Markoviano, os grafos gerados inicialmente com os modelos teóricos perdem suas propriedades básicas iniciais e passam a ter propriedades sem um padrão

definido. Em outras palavras, os valores das métricas de coeficiente de clusterização, distribuição do grau e diâmetro da rede, característicos dos modelos teóricos são perdidos, fazendo com que tais métricas possuam valores aleatórios. Diante disso, o comportamento dos algoritmos de difusão fica semelhante ao comportamento que apresentavam quando aplicados em grafos aleatórios. Tais resultados indicam que, ao utilizar tal gerador, não existe a garantia de que as propriedades dos grafos iniciais sejam mantidas, impedindo assim uma análise do comportamento dos algoritmos propostos.

Quanto maior o nível de dinâmica do sistema, mais próximo ao Epidêmico ficam os demais algoritmos de difusão, chegando ao ponto de todos os algoritmos apresentarem um comportamento equivalente, quando aplicados em um cenário de alta dinâmica. Tais resultados reforçam a intuição de que o tempo de difusão dos algoritmos é inversamente proporcional ao grau de dinâmica dos sistemas.

6.6 UTILIZAÇÃO DAS MÉTRICAS EM UM RAIOS DE ABRANGÊNCIA

Nesta seção serão apresentados os resultados referentes à teoria proposta na Seção 4.3, onde são consideradas as métricas de vizinhos pertencentes a um raio \mathcal{R} na escolha do próximo salto. Conforme apresentado na Seção 4.3, a intuição por trás desta abordagem é que um nó, com um valor de métrica mediano, porém com vizinhos que possuem valores elevados de suas métricas, seja um melhor difusor que outro nó com valor de métrica maior, mas cercado por nós com uma menor centralidade.

Para a geração dos resultados foi realizada uma média ponderada dos valores das métricas pertencentes aos nós da vizinhança, onde o peso de tal média é definido pela função de decaimento $f_d = \frac{1}{r}$, onde r corresponde ao raio do vizinho analisado, conforme apresentado na Seção 4.3. Esta média faz com que os vizinhos mais próximos ao nó que difundirá a informação tenham um peso maior do que os vizinhos mais distantes. Os valores utilizados do raio máximo avaliado \mathcal{R} nos experimentos são apresentados na Tabela 6.2.

Para o SopCast, em relação ao algoritmo *Betweenness Walk*, o ganho foi próximo de 1.1% e permaneceu constante para todos os parâmetros utilizados. Para o algoritmo de difusão *Closeness Walk* o ganho inicial também foi próximo de 1.1%, mas houve um

	Raio \mathcal{R}
SopCast	1, 3, 5
Infocom	1, 3, 5, 10, 15

Tabela 6.2: Valores utilizados de \mathcal{R} .

pequeno crescimento, chegando a alcançar o ganho de 1.5%. Em relação ao algoritmo *Degree Walk*, o ganho médio foi próximo a 4.7%. Diante disso, podemos concluir que no SopCast, um nó tem um maior poder de difusão de informação quando está próximo a outros nós que possuem alto valor do Grau. Isso ocorre pelo fato do diâmetro da rede ser pequeno.

Em relação ao Infocom, para o algoritmo de difusão *Betweenness Walk*, o ganho ficou no intervalo de 0.9% a 3%. Para o algoritmo *Closeness Walk*, o intervalo do ganho foi 0.5% a 1.9%. Já em relação ao algoritmo *Degree Walk*, o ganho foi constante, próximo de 1.3%. A métrica *Betweenness* obteve o seu melhor resultado quando utilizada juntamente com o maior valor do raio utilizado ($\mathcal{R} = 15$), valor corresponde ao diâmetro desta rede, utilizando assim o conhecimento global da rede na avaliação das métricas.

É fácil perceber que existe um aumento no custo computacional do algoritmo, devido a necessidade de realizar o cálculo das métricas em um número maior de nós da rede. A análise do custo desta proposta será realizada da forma como segue. Seja D a densidade da rede, definida como sendo a proporção do o número de ligações da rede no instante atual ($|E|$) pelo número máximo possível de ligações, valor este dado pela binomial $\binom{|V|}{2}$, onde $|V|$ representa o número de entidades da rede. Sendo assim, D será definido pela Equação 6.1. O custo então será proporcional a $D * \mathcal{R}^2 * \mathcal{J}_c$, onde \mathcal{J}_c é o custo do algoritmo original definido na Seção 4.2.1.

$$D = \frac{2 * |E|}{|V| * (|V| - 1)}. \quad (6.1)$$

Diante de tais custos, é necessário analisar se os ganhos alcançados compensam o aumento do custo. O maior ganho alcançado na rede SopCast foi próximo a 5%. Este valor chega a ser bem próximo a 1 segundo no tempo necessário para difundir a informação a toda a rede. Este ganho pode ser satisfatório, pela característica de compartilhamento de vídeo que, na maioria dos casos, é transmitido ao vivo. Já em relação ao Infocom, o maior ganho alcançado foi de 3%, o que corresponde a aproximadamente um ganho 1.2

segundos. Este ganho pode não ser satisfatório, uma vez que a rede é utilizada pelos membros do Infocom principalmente para compartilhamento de arquivos pessoais.

Estes resultados são indícios que é possível alcançar ganhos ainda maiores com relação ao tempo de difusão, utilizando para isso não somente as métricas diretas de seus vizinhos, mas realizando algum tratamento algorítmico de características dos demais membros da rede. Cabe ressaltar que existirá um aumento considerável do custo do algoritmo, que deve ser levado em consideração na implementação e aplicação da proposta.

6.7 NÚMERO DE RÉPLICAS GERADAS NA REDE

Nesta seção são apresentados os ganhos ao se utilizar processos de difusão orientados pela centralidade, em relação ao número de réplicas das mensagens na rede, onde a réplica é definida como sendo uma duplicação da mensagem na retransmissão. Este valor é calculado a cada iteração do algoritmo, onde um contador é incrementado cada vez que um dispositivo repassa uma mensagem a algum de seus vizinhos. As Figuras 6.11(a), 6.11(b) e 6.11(c) apresentam o número de réplicas das mensagens trocadas pelos processos de difusão nas redes SopCast, Infocom e Sintéticas, respectivamente.

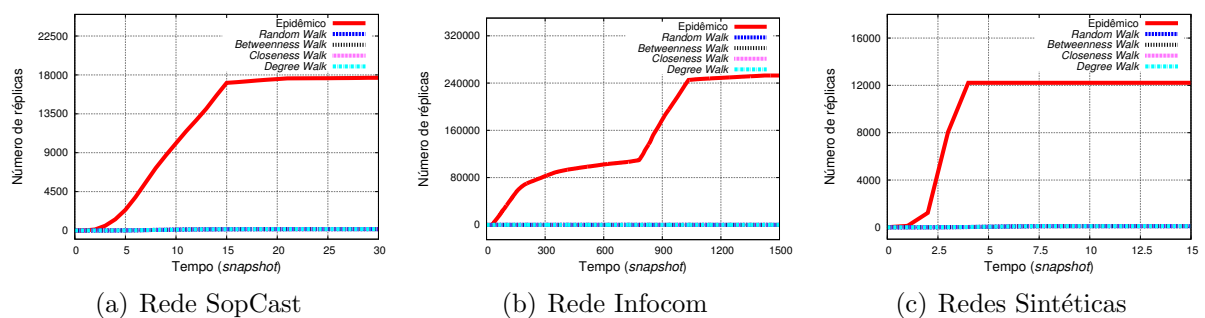


Figura 6.11: Número de réplicas geradas

Como é possível observar, o número de réplicas de mensagens geradas pelos algoritmos propostos neste trabalho foi equivalente ao algoritmo *Random Walk* em ambas as redes. Além disso, estes valores estão em escalas bem menores que o número de mensagens trocadas pelo algoritmo Epidêmico. Por exemplo, nas redes SopCast e Sintéticas os algoritmos orientados por centralidade estão a três ordens de grandeza do algoritmo Epidêmico. Já em relação à rede Infocom, a ordem de grandeza é quatro vezes menor.

Estes resultados reforçam a ideia de que os algoritmos *Betweenness Walk*, *Closeness*

Walk e *Degree Walk* não somente diminuem o tempo de difusão de informação de uma rede dinâmica, mas diminuem também a sobrecarga do sistema, baseado no número de réplicas geradas pelos algoritmos de difusão. Cabe ressaltar que o custo computacional do cálculo dos valores das métricas não foi considerado nesta afirmação.

6.8 PREVISÃO DE *LINKS*

Nesta seção serão apresentados os resultados obtidos aplicando a teoria de previsão de *links* apresentada na Seção 4.4.

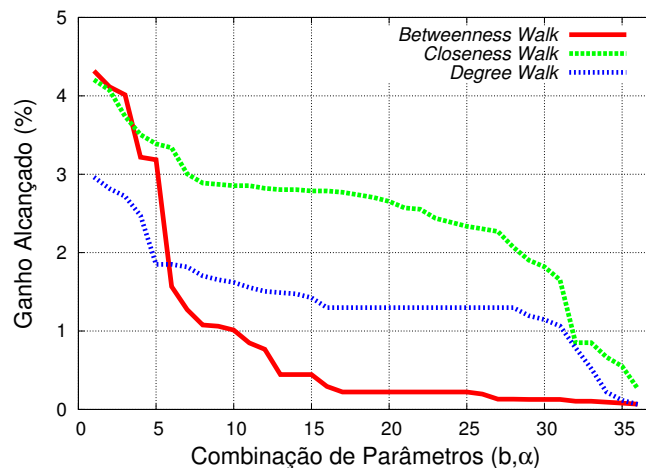
Os resultados apresentados nesta seção consideram a variação dos parâmetros b e α . Conforme definido na Seção 4.4, o parâmetro b indica o número de grafos do histórico que será avaliado na previsão e o parâmetro α ($0 < \alpha < 1$) corresponde ao peso de cada aresta no cálculo do *Weighted Moving Average*. Claramente, a escolha da combinação dos valores de b e α influencia a acurácia da previsão. Os resultados são obtidos considerando combinações de b e α , de tal forma que b varie de 5 a 50 e α varie de 0.1 a 0.9. Aos parâmetros λ e ρ presentes na Cadeia de Markov, foram atribuídos os valores 0.5 em ambos. Estes parâmetros definem se o estado da aresta será modificado pela previsão. Tal valor implica que, se a previsão indicar uma probabilidade de alteração no estado maior que 50%, o algoritmo mudará o estado atual da aresta.

As Figuras 6.12(a) e 6.12(b) mostram o impacto da previsão na difusão baseada em centralidade nos conjuntos de dados Infocom e SopCast. Tais figuras mostram os ganhos alcançados, calculados da forma apresentada na Equação 6.2. São comparados os resultados da previsão de aresta com os valores encontrados usando a mesma métrica de centralidade, sem realizar previsão, para diferentes as amostras combinadas de b e α , nos intervalos apresentados. Estas amostras combinadas de b e α são ordenadas no eixo x . Para ambos os conjuntos de dados, é possível notar claramente que a previsão das conexões do próximo grafo acelera o processo de difusão. Por exemplo, usando o a centralidade *betweenness*, a aceleração do processo de difusão de dados atinge o valor de 4.3%. Para o conjunto de dados SopCast, observa-se um fator de aceleração média próximo a 11.6% quando é utilizada a difusão baseada no Grau. Este resultado foi alcançado devido a estrutura do grafo do SopCast, com o diâmetro próximo a 4, permitindo assim que nós

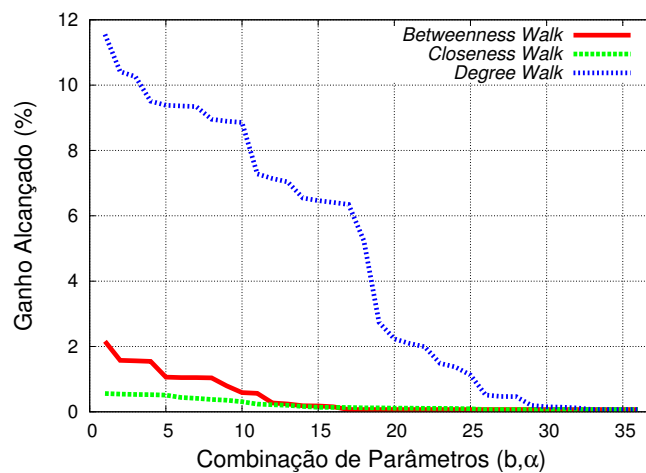
com o maior grau no instante futuro difunda a mensagem mais rapidamente.

$$\text{Ganho Alcançado} = \left(1 - \frac{\text{difusão com previsão}}{\text{difusão sem previsão}} \right) * 100. \quad (6.2)$$

Tais resultados demonstram que, como era esperado, é possível ter uma diminuição no tempo de difusão de uma rede se a mensagem for repassada ao vizinho com maior poder de difusão no próximo instante de tempo, instante esse que tal vizinho encaminhará a mensagem.



(a) Log Infocom



(b) Log SopCast

Figura 6.12: Aceleração da difusão de informação usando a previsão de arestas.

6.9 CONSIDERAÇÕES

Neste capítulo foram apresentados os resultados alcançados nesta dissertação. Ao desconsiderar a dinâmica inerente dos sistemas, o tempo de difusão de uma informação pela

rede é subestimado. É possível acelerar o tempo de difusão de informação na rede, utilizando para isso medidas de centralidade dos nós. Foi apresentada também a diminuição do tempo de difusão dos algoritmos, quando aplicado um modelo simples de previsão de arestas.

7 CONCLUSÃO

Neste trabalho é investigado o impacto gerado ao se adotar métricas de centralidade no processo de difusão de informação em redes complexas dinâmicas. São apresentadas comparações dos desempenhos alcançados dos algoritmos de difusão quando a dinâmica do sistema é considerada e também quando é descartada. Se a dinâmica da rede é desconsiderada, um modelo típico de processo de difusão pode subestimar significativamente o tempo de cobertura de rede.

O conceito de dinâmica é amplo e sua modelagem claramente influencia o desempenho dos algoritmos de difusão de informação. Esta afirmação é baseada nos resultados onde são aplicados os algoritmos de difusão propostos, tanto nos cenários reais quanto nos sintéticos.

A utilização das características topológicas dos nós que formam a rede no processo de cobertura aumenta o desempenho dos modelos de difusão. Uma seleção do próximo salto baseada em métricas de centralidade pode acelerar o processo de difusão em redes dinâmicas com um custo relativamente baixo, tanto no número de mensagens quanto no volume de tráfego gerado. Conforme observado, é possível diminuir o tempo para difundir a informação a uma grande parcela dos destinos quando se considera as métricas topológicas de *Betweenness*, *Closeness* ou até mesmo o Grau, que é uma métrica onde sua obtenção possui um custo significativamente menor que as anteriores citadas.

Neste trabalho é apresentado também resultados sobre maneiras de acelerar ainda mais o processo de difusão, reduzindo assim o tempo de cobertura de rede. É possível melhorar a difusão por métricas, utilizando para isso valores das métricas da vizinhança pertencente a um raio de abrangência, além da adoção de um sistema de previsão das arestas presentes no próximo instante de tempo.

Estes resultados são encorajadores para o prosseguimento da investigação sobre a previsão como um fator de aceleração de processos de difusão movidos por centralidade em redes complexas dinâmicas. Este trabalho pode ainda ser estendido com as seguintes linhas de atuação:

- Investigação sobre técnicas que determinam combinações convenientes dos parâmetros envolvidos.

- Estudo de diferentes formas de prever a topologia da rede nos instantes futuros, como por exemplo a utilização de um previsor baseado em redes neurais.
- Embasamento teórico dos resultados encontrados e ganhos alcançados.

REFERÊNCIAS

- ACER, U. G.; DRINEAS, P.; ABOUZEID, A. A. Connectivity in time-graphs. **Pervasive and Mobile Computing**, v. 7, n. 2, p. 160 – 171, 2011. ISSN 1574-1192. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574119210001288>>.
- ADAMIC, L. A.; ADAR, E. How to search a social network. **Social Networks**, v. 27, p. 187–203, 2005.
- ADAMIC, L. A.; LUKOSE, R. M.; PUNIYANI, A. R.; HUBERMAN, B. A. Search in power-law networks. **Physical Review E**, American Physical Society, v. 64, n. 4, p. 046135+, set. 2001. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.64.046135>>.
- ALVAREZ-HAMELIN, J. I.; FLEURY, E.; VESPIGNANI, A.; ZIVIANI, A. Complex dynamic networks: Tools and methods (Guest Editorial). **Computer Networks**, v. 56, n. 3, p. 967–969, dez. 2012. ISSN 13891286.
- ASH, G. R. **Dynamic Routing in Telecommunications Networks**. 1st. ed., 1997. ISBN 0070064148.
- BALACHANDRAN, A.; VOELKER, G.; BAHL, P.; RANGAN, P. Characterizing user behavior and network performance in a public wireless lan. In: ACM. **ACM SIGMETRICS Performance Evaluation Review**, 2002. v. 30, n. 1, p. 195–205.
- BARABÁSI, A.-L.; BONABEAU, E. Scale-free networks. **Scientific American**, v. 288, p. 50–59, 2003.
- BARABÁSI, A.-L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509–512, 1999. Disponível em: <<http://www.sciencemag.org/content/286/5439/509.abstract>>.
- BARBERA, M.; STEFA, J.; VIANA, A.; AMORIM, M. de; BOC, M. Vip delegation: Enabling vips to offload data in wireless social mobile networks. In: **Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on**, 2011. p. 1 –8.

- BASU, P.; BAR-NOY, A.; RAMANATHAN, R.; JOHNSON, M. P. Modeling and Analysis of Time-Varying Graphs. Published on arXiv.org;cs;arXiv:1012.0260. dez. 2010.
- BLUETOOTH, S. Specification of the bluetooth system, version 1.1. <http://www.bluetooth.com>, 2001. Disponível em: <<http://ci.nii.ac.jp/naid/10012296686/en/>>.
- BOCCALETTI, S.; LATORA, V.; MORENO, Y.; CHAVEZ, M.; HWANG, D. Complex networks: Structure and dynamics. **Physics Reports**, v. 424, p. 175–308, 2006.
- BOGUÑÁ, M.; KRIOUKOV, D.; CLAFFY, k. Navigability of Complex Networks. **Nature Physics**, v. 5, n. 1, p. 74–80, 2009.
- BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. In: **Proceedings of the seventh international conference on World Wide Web 7**, 1998. (WWW7), p. 107–117. Disponível em: <<http://dl.acm.org/citation.cfm?id=297805.297827>>.
- CALDARELLI, G.; VESPIGNANI, A. **Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science (Complex Systems and Interdisciplinary Science)**, 2007. Hardcover. ISBN 981270664X. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/981270664X>>.
- CARLEY, K. M. Dynamic network analysis. In: COMMITTEE ON HUMAN FACTORS, NATIONAL RESEARCH COUNCIL. **Dynamic social network modeling and analysis: Workshop summary and papers**, 2003. p. 133–145.
- CARRINGTON, P. J.; SCOTT, J.; WASSERMAN, S. **Models and methods in social network analysis**. Cambridge university press, 2005.
- CASE, J.; FEDOR, M.; SCHOFFSTALL, M.; DAVIN, C. **A simple network management protocol (SNMP)**. Network information center, sri international, 1989.
- CHARENTREAU, A.; HUI, P.; CROWCROFT, J.; DIOT, C.; GASS, R.; SCOTT, J. Impact of Human Mobility on Opportunistic Forwarding Algorithms. **IEEE Trans. Mob. Comp**, v. 6, p. 606–620, 2007.

- COOLEN, A.; MARTINO, A.; ANNIBALE, A. Constrained markovian dynamics of random graphs. **Journal of Statistical Physics**, Springer US, v. 136, p. 1035–1067, 2009. ISSN 0022-4715. Disponível em: <<http://dx.doi.org/10.1007/s10955-009-9821-2>>.
- DEMERS, A.; GREENE, D.; HAUSER, C.; IRISH, W.; LARSON, J.; SHENKER, S.; STURGIS, H.; SWINEHART, D.; TERRY, D. Epidemic algorithms for replicated database maintenance. In: **Proceedings of the sixth annual ACM Symposium on Principles of distributed computing**, 1987. (PODC '87), p. 1–12. ISBN 0-89791-239-X. Disponível em: <<http://doi.acm.org/10.1145/41840.41841>>.
- DEZSO, Z. Diffusion on networks. In: **APS Meeting Abstracts**, 2004. v. 1, p. 18001.
- DOROGOVTSEV, S. N.; MENDES, J. F. F. **Evolution of Networks: From Biological Nets to the Internet and WWW (Physics)**. Oxford university press, inc., 2003. ISBN 0198515901.
- EASLEY, D.; KLEINBERG, J. **Networks, Crowds, and Markets: Reasoning about a Highly Connected World**. Cambridge university press, 2010.
- EICHEL, P.; DELP, E. Sequential edge linking. In: **Proceedings 22nd Allerton Conference on Communication, Control, and Computation**, 1984. p. 782–791.
- ERDÖS, P.; RÉNYI, A. On Random Graphs, I. **Publicationes Mathematicae**, v. 6, p. 290–297, 1959.
- FACEBOOK. <http://www.facebook.com/>. 2013.
- FALL, K.; FARRELL, S. Dtn: an architectural retrospective. **Selected Areas in Communications, IEEE Journal on**, v. 26, n. 5, p. 828 –836, june 2008. ISSN 0733-8716.
- FEIT, S. M. **SNMP: A Guide to Network Management**, 1993. ISBN 0070203598.
- FERREIRA, F. H. C.; VIEIRA, A. B.; SILVA, A. P. C. **Caracterização da dinâmica de participantes e Comunidades em sistemas de P2P de transmissão de vídeo ao vivo**. Dissertação (Mestrado) — PGCC, Programa de Pós-Graduação em Ciência da Computação, UFJF, 2013.

- FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, American Sociological Association, v. 40, n. 1, p. pp. 35–41, 1977. ISSN 00380431. Disponível em: <<http://www.jstor.org/stable/3033543>>.
- GERLA, M.; TANG, K.; BAGRODIA, R. Tcp performance in wireless multi-hop networks. In: **Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA '99. Second IEEE Workshop on**, 1999. p. 41–50.
- GONÇALVES, K.; VIEIRA, A.; ALMEIDA, J.; SILVA, A. da; MARQUES-NETO, H.; CAMPOS, S. Characterizing dynamic properties of the sopcast overlay network. In: **IEEE. Parallel, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on**, 2012. p. 319–326.
- GOOGLE. <http://www.google.com/>. 2013.
- HAGBERG, A. A.; SCHULT, D. A.; SWART, P. J. Exploring network structure, dynamics, and function using NetworkX. In: **Proceedings of the 7th Python in Science Conference (SciPy2008)**, 2008. p. 11–15.
- HUI, P.; CROWCROFT, J.; YONEKI, E. Bubble rap: Social-based forwarding in delay-tolerant networks. **Mobile Computing, IEEE Transactions on**, v. 10, n. 11, p. 1576–1589, nov. 2011. ISSN 1536-1233.
- INTANAGONWIWAT, C.; GOVINDAN, R.; ESTRIN, D. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In: **Proceedings of the 6th annual international conference on Mobile computing and networking**, 2000. (MobiCom '00), p. 56–67. ISBN 1-58113-197-6. Disponível em: <<http://doi.acm.org/10.1145/345910.345920>>.
- JE, C. Infectious diseases of humans: Dynamics and control. **JAMA**, v. 268, n. 23, p. 3381, 1992. Disponível em: <+<http://dx.doi.org/10.1001/jama.1992.03490230111047>>.
- KIM, B. J.; YOON, C. N.; HAN, S. K.; JEONG, H. Path finding strategies in scale-free networks. **Physical Review E**, v. 65, p. 027103, Jan 2002.

- KITSAK, M.; GALLOS, L.; HAVLIN, S.; LILJEROS, F.; MUCHNIK, L.; STANLEY, H.; MAKSE, H. Identification of influential spreaders in complex networks . **Nature Physics**, v. 6, n. 11, p. 888–893, Aug 2010.
- KOCAREV, L. Network Science: A New Paradigm Shift. **IEEE Network**, v. 24, n. 6, p. 6–9, 2010.
- KOTZ, D.; ESSIEN, K. Analysis of a campus-wide wireless network. In: **Proceedings of the 8th annual international conference on Mobile computing and networking**, 2002. (MobiCom '02), p. 107–118. ISBN 1-58113-486-X. Disponível em: <<http://doi.acm.org/10.1145/570645.570659>>.
- KOTZ, D.; HENDERSON, T.; ABYZOV, I.; YEO, J. **CRAWDAD data set dartmouth/campus (v. 2009-09-09)**. set. 2009. Downloaded from <http://crawdad.cs.dartmouth.edu/dartmouth/campus>.
- LINDGREN, A.; DORIA, A.; SCHELÉN, O. Probabilistic routing in intermittently connected networks. **SIGMOBILE Mob. Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 7, n. 3, p. 19–20, jul. 2003. ISSN 1559-1662. Disponível em: <<http://doi.acm.org/10.1145/961268.961272>>.
- LOVASZ, L. Combinatorics, Paul Erdos is Eighty. **Bolyai Society Mathematical Studies**, v. 2, p. 1–46, 1993.
- MTIBAA, A.; MAY, M.; DIOT, C.; AMMAR, M. Peoplerank: Social opportunistic forwarding. In: **INFOCOM, 2010 Proceedings IEEE**, 2010. p. 1 –5. ISSN 0743-166X.
- MYSPEACE. <http://www.myspace.com/>. 2013.
- NES, J. G. G.; LATORA, V. Entropy rate of diffusion processes on complex networks. **Physical Review E**, American Physical Society, v. 78, n. 6, p. 065102+, dez. 2008. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.78.065102>>.
- NEWMAN, M. **Networks: An Introduction**, 2010. ISBN 0199206651, 9780199206650.
- NEWMAN, M. E. The structure and function of complex networks. **SIAM review**, SIAM, v. 45, n. 2, p. 167–256, 2003.

OKAMOTO, K.; CHEN, W.; LI, X.-Y. Ranking of closeness centrality for large-scale social networks. In: PREPARATA, F.; WU, X.; YIN, J. (Ed.). **Frontiers in Algorithms**, 2008, (Lecture Notes in Computer Science, v. 5059). p. 186–195. ISBN 978-3-540-69310-9.

ORKUT. <http://www.orkut.com/>. 2013.

PIETILÄNEN, A.-K.; DIOT, C. Dissemination in opportunistic social networks: the role of temporal communities. In: **Proceedings of the thirteenth ACM international symposium on Mobile Ad Hoc Networking and Computing**, 2012. (MobiHoc '12), p. 165–174. ISBN 978-1-4503-1281-3. Disponível em: <<http://doi.acm.org/10.1145/2248371.2248396>>.

RANGANATHAN, K.; IAMNITCHI, A.; FOSTER, I. Improving data availability through dynamic model-driven replication in large peer-to-peer communities. In: **Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on**, 2002. p. 376–376.

RANGO, F. D.; MONTEVERDI, F. Social and dynamic graph-based scalable routing protocol in a dtn network. In: **Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 2012 International Symposium on**, 2012. p. 1–8.

ROBERTS, S. W. Control chart tests based on geometric moving averages. **Technometrics**, v. 1, n. 3, p. 239–250, 1959. Disponível em: <<http://amstat.tandfonline.com/doi/abs/10.1080/00401706.1959.10489860>>.

ROSVALL, M.; GRONLUND, A.; MINNHAGEN, P.; SNEPPEN, K. Searchability of networks. **Physical Review E**, v. 72, n. 4, Oct 2005.

SCOTT, J.; GASS, R.; CROWCROFT, J.; HUI, P.; DIOT, C.; CHAINTREAU, A. **CRAWDAD data set cambridge/haggle (v. 2009-05-29)**. maio 2009. Downloaded from <http://crawdad.cs.dartmouth.edu/cambridge/haggle>.

SIMSEK, Ö.; JENSEN, D. Navigating networks by using homophily and degree. **Proceedings of the National Academy of Sciences**, v. 105, n. 35, p. 12758–12762, 2008.

SOPCAST. <http://www.sopcast.org/>. 2013.

TANG, S.; LU, Y.; HERNÁNDEZ, J. M.; KUIPERS, F.; MIEGHEM, P. Topology dynamics in a p2ptv network. In: **Proceedings of the 8th International IFIP-TC 6 Networking Conference**, 2009. (NETWORKING '09), p. 326–337. ISBN 978-3-642-01398-0.

TARIQ, M. M. B.; AMMAR, M.; ZEGURA, E. Message ferry route design for sparse ad hoc networks with mobile nodes. In: **Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing**, 2006. (MobiHoc '06), p. 37–48. ISBN 1-59593-368-9. Disponible em: <<http://doi.acm.org/10.1145/1132905.1132910>>.

UGANDER, J. Scaling graph computations at facebook. In: **Proceedings of the 22nd international conference on World Wide Web companion**, 2013. (WWW '13 Companion), p. 499–500. ISBN 978-1-4503-2038-2. Disponible em: <<http://dl.acm.org/citation.cfm?id=2487788.2487982>>.

VAHDAT, A.; BECKER, D. **Epidemic routing for partially-connected ad hoc networks**, Technical Report CS-200006, Duke University, 2000.

VALENTE, T.; CORONGES, K.; LAKON, C.; COSTENBADER, E. How correlated are network centrality measures? **Connections**, NIH Public Access, v. 28, n. 1, p. 16–26, 2008.

VALENTE, T. W. Network models of the diffusion of innovations. **Computational & Mathematical Organization Theory**, Springer, v. 2, n. 2, p. 163–164, 1996.

VOGELS, W.; RENESSE, R. van; BIRMAN, K. The power of epidemics: robust communication for large-scale distributed systems. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 33, n. 1, p. 131–135, jan. 2003. ISSN 0146-4833. Disponible em: <<http://doi.acm.org/10.1145/774763.774784>>.

WANG, L.; HOPCROFT, J.; HE, J.; LIANG, H.; SUWAJANAKORN, S. Extracting the core structure of social networks using (a, b)-communities. **Internet Mathematics**, v. 9, n. 1, p. 58–81, 2013. Disponible em: <<http://www.tandfonline.com/doi/abs/10.1080/15427951.2012.678187>>.

- WANG, X. F. Complex networks: topology, dynamics and synchronization. **International Journal of Bifurcation and Chaos**, World Scientific, v. 12, n. 05, p. 885–916, 2002.
- WATANABE, M.; SUZUMURA, T. How social network is evolving?: a preliminary study on billion-scale twitter network. In: **Proceedings of the 22nd international conference on World Wide Web companion**, 2013. (WWW '13 Companion), p. 531–534. ISBN 978-1-4503-2038-2. Disponível em: <<http://dl.acm.org/citation.cfm?id=2487788.2487988>>.
- WATTS, A. A dynamic model of network formation. In: **Networks and Groups**, 2003. p. 337–345.
- WATTS, D.; STROGATZ, S. Collective dynamics of 'small-world' networks. **Nature**, v. 393, n. 6684, p. 440–442, jun. 1998. Disponível em: <<http://europemc.org/abstract/MED/9623998>>.
- WEHMUTH, K.; ZIVIANI, A. Um Novo Algoritmo Distribuido para Avaliação e Localização de Centralidade de Rede. In: **Proceedings of X Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance)**, 2011.
- YOSHIDA, N. Estimation for diffusion processes from discrete observation. **Journal of Multivariate Analysis**, v. 41, n. 2, p. 220–242, 1992. ISSN 0047-259X.
- YUAN, P.; MA, H. Hug: Human gathering point based routing for opportunistic networks. In: **Wireless Communications and Networking Conference (WCNC), 2012 IEEE**, 2012. p. 3024 –3029. ISSN 1525-3511.