

# Análise da Disseminação de Informação em Redes de Comunicação Considerando Dinâmica

Abraão Guimarães Flores<sup>1</sup>, Alex Borges<sup>1</sup>, Ana Paula Couto da Silva<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação  
Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal de Juiz de Fora (UFJF)  
CEP 36.036-900 – Juiz de Fora – MG – Brazil

abraao.guimaraes@ice.ufjf.br  
{alex.borges, anapaula.silva}@ufjf.edu.br

**Abstract.** *Dynamic is a common feature present in many real systems such as communication, social, biological and technological networks. When evaluating a particular real system, including the dynamic aspect of it, we are faced with two questions: first, how to define the dynamics itself; second, how to capture changes in such real systems characteristics, avoiding a large amount of information in a single instant of time. The contributions of this paper are two-fold. First we discuss the impact dynamics in the real systems modeling. We show that the predicted time to disseminate information on the network do not correspond to reality when calculated without taking into account the dynamics of the system. As a second key contribution, we check the gain when considering topological metric dynamics to accelerate the information dissemination process in these systems. We simulate 3 information diffusion algorithms on top of real networks topology. We show that the SoA algorithm achieves up to 20% better performance than Random Walk. Moreover, the CuT algorithm can achieve the minimum transmission time, given by an epidemic spreading model.*

**Resumo.** *A dinamicidade é uma característica constante em diversos sistemas reais, tais como redes de comunicação, sociais, biológicas e tecnológicas. Ao avaliarmos um determinado sistema real, incluindo os aspectos dinâmicos do mesmo, nos deparamos com duas questões. Primeiro, como modelar a dinâmica propriamente dita. Em segundo, como capturar mudanças das características de tais sistemas, evitando agregar uma grande quantidade de informação em um único instante de tempo. As contribuições deste artigo se dividem em duas vertentes principais. Primeiramente é abordado o impacto em considerar a dinâmica na modelagem de sistemas reais. Nesse caso, os valores do tempo de difusão da informação na rede não correspondem à realidade quando preditos sem levar em conta a dinâmica do sistema. A segunda contribuição é a verificação do ganho ao se considerar métricas topológicas dinâmicas para acelerar o processo de difusão de informação nestes sistemas. São simulados 3 algoritmos de difusão de informação em cima de topologias reais de rede. No caso de se utilizar o modelo SoA, há um ganho de até 20% em relação ao Random Walk para se alcançar o mesmo percentual de usuários. Ao se utilizar o CuT, se alcança o tempo de difusão mínimo da rede, dado pelo modelo epidêmico de difusão.*

## 1. Introdução

A presença do conceito de redes como conjunto de entidades que possuem ligações entre si está presente em diversas áreas da vida cotidiana. Entre os diversos exemplos destes tipos de redes, estão presentes as redes de comunicação, sociais, biológicas e tecnológicas.

Uma característica em comum entre estes sistemas é o dinamismo. 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. 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. Como consequência, ignorar o dinamismo pode comprometer propostas mais eficazes de gerência, construção e caracterização das redes encontradas na vida real.

Poucos trabalhos na literatura consideram a dinâmica do sistema na caracterização do mesmo. Isso ocorre principalmente devido à alta complexidade introduzida tanto na modelagem quanto na análise das características comportamentais de maior relevância. O próprio conceito de dinâmica em um sistema real não é trivial, e proposições devem ser estabelecidas *a priori* da análise a ser realizada.

A difusão de informação entre os elementos da rede é um dos diversos problemas importantes que devem ser abordados no estudo de redes reais, quando se considera a dinâmica. 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 qualidades dos integrantes das mesmas. 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. Neste caso em específico, se o objetivo é que uma informação alcance um maior número de indivíduos em um menor intervalo de tempo, indivíduos com maior número de contatos devem ser escolhidos para receber a informação a ser disseminada. Este é um cenário simples onde a dinâmica da rede, representada pela inclusão/supressão de contatos, influencia na escolha da entidade que receberá a informação a ser propagada.

Diversas redes possuem um grau de dinamicidade elevado, e assim, suas características se modificam em um curto intervalo de tempo. Ao se analisar estas redes após um longo intervalo de observações, pode-se perder informações valiosas para uma análise detalhada do comportamento destes sistemas. Este fenômeno ocorre por conta de se agregar, em uma única fotografia, um conjunto grande de modificações estruturais. Tais modificações, em muitos casos, deveriam ser verificadas em intervalos de tempo menores.

Assim, ao se avaliar um determinado sistema real, incluindo o aspecto dinâmico do mesmo, deve-se considerar duas questões importantes. Primeiro, como definir a *dinâmica* propriamente dita. Em segundo, como capturar mudanças de suas características evitando agregar uma grande quantidade de informação.

Em relação a primeira questão, neste artigo define-se dinâmica considerando três diferentes aspectos. O primeiro refere-se à mobilidade das entidades no sentido de mudanças ocorridas em suas listas de contatos. O segundo refere-se à mudança de características topológicas do grafo que representa o sistema estudado, especialmente características de centralidade, que modelam os sistemas reais. E por último, considera-se o nível de agregação de informação que será analisada, sendo esta proporcional ao intervalo de tempo utilizado para fornecer as fotografias que caracterizam os estados do sistema.

As contribuições deste artigo se dividem em duas vertentes principais. Primeiro, é abordado o impacto em considerar as diferentes definições de dinâmica na modelagem de sistemas reais. É considerado o problema de difusão de informação em dois cenários de redes dinâmicas: uma rede *wireless* de um *campus* universitário americano e a aplicação P2P SopCast [Sopcast 2012]. Segundo, verifica-se o ganho ao se considerar métricas topológicas dinâmicas para acelerar o processo de difusão de informação nestes sistemas.

Os resultados encontrados mostram que, introduzir a dinâmica na modelagem dos sistemas em estudo fornece conclusões mais próximas do comportamento real das entidades pertencentes a estes sistemas. De fato, quando não se considera a dinâmica de tais sistemas, os valores encontrados para o tempo de difusão da informação na rede não correspondem à realidade. Entre os 3 algoritmos de difusão de informação simulados em cima de topologias reais de rede, o SoA apresenta um ganho de 20% em relação ao *Random Walk*. Ao se utilizar o CuT, se alcança o tempo de difusão mínimo da rede, dado pelo modelo *epidêmico* de difusão.

O restante desse artigo está organizado como segue. A Seção 2 apresenta os trabalhos relacionados ao presente estudo. Os conceitos teóricos que fundamentam a análise, incluindo a dinâmica de redes de comunicação, são apresentados na Seção 3. Na Seção 4 são apresentados modelos de difusão que consideram métricas topológicas para acelerar a difusão de informação na rede. Os *logs* das redes de comunicação analisadas neste artigo são descritos na Seção 5. Resultados numéricos são apresentados na Seção 6. Finalmente, a Seção 7 discute as principais conclusões e os trabalhos futuros.

## 2. 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.

Em [Barbera et al. 2011], os autores tratam do problema de difusão de informação aplicada a diferentes tipos de redes. O principal objetivo é cobrir o maior percentual possível de nós integrantes no sistema, utilizando um conjunto de usuários importantes, denominados *VIPs*. A movimentação destes usuários permite o estabelecimento de contatos regulares com os demais usuários da rede. A classificação de um usuário como *VIP* é baseada em métricas topológicas do grafo (estático) que modela o sistema. A análise apresentada neste artigo difere da análise apresentada em [Barbera et al. 2011] em dois pontos: (i) o modo como a informação é difundida entre os nós e; (ii) a métrica de interesse avaliada. Em [Barbera et al. 2011], os nós classificados como *VIPs* circulam entre os demais nós e a métrica de interesse é o percentual de nós que receberam a informa-

ção. Neste artigo, a entrega da informação é *multi-hop* e a métrica de interesse é o tempo necessário para entrega da informação.

Em [Hui et al. 2008], os autores consideram métricas mais elaboradas para a escolha da entidade a receber a informação inicial. O objetivo é diminuir o tempo de propagação entre os diversos participantes. A escolha do nó a receber a informação é baseada nos conceitos de centralidade e comunidade. No entanto, os autores não consideram a dinâmica do sistema nos cálculos destas métricas.

A importância da identificação das entidades que potencialmente possam difundir mais rapidamente a informação dentro de uma rede também é explorada em [Kitsak et al. 2010]. São utilizados dados de duas redes no estudo: a rede criada pelos contatos entre pacientes em hospitais da Suécia e redes de atores que estrelaram filmes em comum. Segundo os autores, as entidades mais indicadas para receber a informação são as localizadas no núcleo da rede. Os diferentes extratos das redes são definidos através da teoria de *k-shell decomposition analysis*. Diferentemente dos resultados apresentados neste artigo, os resultados de [Kitsak et al. 2010] não consideram a dinâmica, ou seja, possuem uma única fotografia que representa todo o tempo de observação.

### 3. Referencial Teórico

#### 3.1. Modelagem Matemática

A notação e a definição matemática para modelar as redes dinâmicas utilizadas nesse trabalho são baseadas em [Basu et al. 2010].

Uma rede é 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. Seja o período total de observação do sistema real denotado por  $T_N$ , iniciado no tempo  $T_1 = 0$ . Sem perda de generalidade, considera-se que a visão agregada da rede, em outras palavras a modelagem do sistema considerando o tempo total de observação  $T_N$ , é feita através da construção do grafo estático  $G$ .

Seja  $\mathcal{G}_t(\mathcal{V}_t, \mathcal{E}_t)$  a representação dinâmica do grafo agregado  $G$ , sendo  $\mathcal{V}_t$  o conjunto de nós e  $\mathcal{E}_t$  o conjunto de arestas. 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 *snapshots* incrementais no tempo.

A duração de cada *snapshot* é definida por  $\Delta = T_i - T_{i-1}$ , com  $0 \leq i \leq N$ . Esta representação permite capturar a evolução da rede no espaço e no tempo. Ou seja, mudanças estruturais são capturadas a cada  $\Delta$  unidades de tempo. Claramente, se  $\Delta = T_N$ , é representada uma única fotografia do sistema, desconsiderando toda a dinâmica, e assim, 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 representando as mudanças estruturais ocorridas em um *snapshot* em particular.

A escolha do parâmetro  $\Delta$  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,  $\Delta$  pode ser definido como o menor intervalo de acontecimento de um evento no sistema (e.g. adição/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  $\Delta$ , maior é o nível de agregação de informação incorporada ao grafo e menor a percepção da dinâmica associada.

### 3.2. Métricas Topológicas

Métricas topológicas são definidas como medidas baseadas em atributos estruturais de um grafo. Estas métricas podem considerar cada participante (nó) em específico ou a visão global do grafo. Neste artigo são utilizadas as métricas relacionadas à caracterização dos nós, conhecidas como métricas de centralidade.

Sem perda de generalidade, será considerado o grafo estático  $G$  para formalização das métricas. Para o conjunto de grafos indexados em  $t$ , com  $T_1 \leq t \leq T_N$ , as métricas são definidas em cada *snapshot*. As principais métricas consideradas neste artigo são:

(1) **Grau**: O grau de um nó  $v$  é definido como o total de arestas incidentes a este nó.

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

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

(3) **Closeness**: A métrica de *closeness* de um nó  $v$  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 de *closeness* é dada por:

$$Closeness(v) = \sum_{\forall v \neq j, j \in V} l(v, j)^{-1}.$$

Para os algoritmos de difusão baseados em métricas de centralidade, são calculadas as métricas através da definição clássica das mesmas, apresentadas nos itens 1, 2 e 3. No entanto, diversos trabalhos na literatura buscam aproximações para os seus valores, visando a diminuição do custo computacional [Wehmuth and Ziviani 2011]. Como trabalhos futuros, para avaliação de redes com maior quantidade de nós, poderão ser aplicados os resultados aproximados propostos na literatura.

### 3.3. Modelos de Vizinhança *Store-or-Advance* e *Cut-Through*

Conforme apresentado em [Basu et al. 2010], os modelos SoA e CuT definem o conjunto de vizinhos de cada um dos nós do sistema. 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 *snapshot* analisado. 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 1 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 6 nós. Em contrapartida, para o mesmo intervalo de tempo, o modelo de difusão CuT alcança 9 nós. A maior velocidade na difusão do conteúdo deve-se ao fato do modelo CuT permitir que sejam alcançados todos os nós de toda a componente ligada no instante de tempo que está sendo analisado.

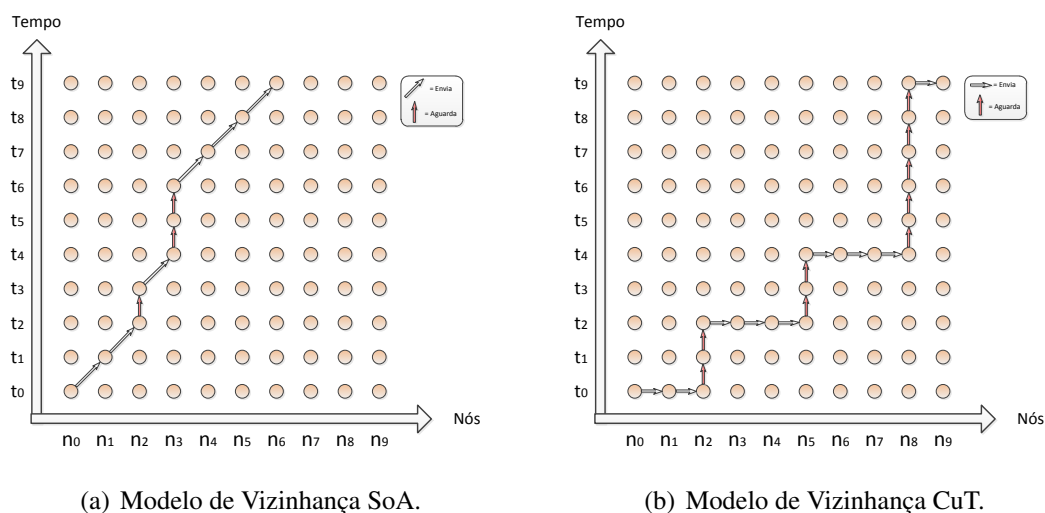


Figura 1. Diferença entre os Modelos de Vizinhaça SoA e CuT.

### 3.4. Modelos de Difusão de Informação

Nesta seção são descritos os principais modelos básicos de difusão de informação, sendo que estes podem ser aplicados tanto no caso onde o sistema é modelado com múltiplas fotografias, ou seja, com o conjunto de grafos  $\mathcal{G}_t$ ,  $T_1 \leq t \leq T_N$ , quanto no caso onde o sistema é modelado através da visão agregada representada pelo grafo  $G$ . Será apresentado o modelo de difusão *Epidêmico* e, baseando-se em [Lovasz 1993], o modelo de difusão *Random Walk* (RW).

#### 3.4.1. *Epidêmico*

A denotação *Epidêmico*, adotada nesse artigo, refere-se à implementação onde todos os vizinhos habilitados recebem a informação a cada iteração, para ambos os modelos SoA e CuT. 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 *snapshot* atual. No entanto, o custo computacional deste método, baseado no número total de mensagens trocadas entre os nós, é elevado. Como consequência, a implementação deste modelo em sistemas reais é complexa e pode-se tornar inviável.

#### 3.4.2. *Random Walk*

Conforme descrito em [Lovasz 1993], o funcionamento do modelo de difusão *Random Walk* (RW) 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 a dinâmica do modelo RW, espera-se que o tempo de difusão da informação seja maior que nos modelos epidêmicos. Adicionalmente, este algoritmo não considera características do sistema real que podem acelerar a difusão da informação entre os nós do grafo. 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.

Nos resultados apresentados na Seção 6, são consideradas duas implementações diferentes do modelo RW: a tradicional, como descrita anteriormente, que pode ser vista como um caso particular do modelo SoA onde a cada instante de tempo somente um vizinho é escolhido para receber a informação que será denotada somente por RW; 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, irá selecionar aleatoriamente um nó dentro da sua vizinhança que receberá a informação a ser difundida.

#### 4. Modelos de Difusão utilizando Métricas Topológicas

Com o objetivo de acelerar o processo de difusão de informação em redes que representam sistemas reais considera-se o conhecimento de características das entidades que formam o sistema real sendo analisado para decidir qual será a próxima entidade a receber a informação a ser difundida no sistema.

Os modelos de difusão, denominados *Betweenness Walk* e *Closeness Walk* se assemelham ao modelo RW, nas implementações SoA e CuT. Seja o grafo  $G$  ou  $G_t$  que modela o sistema real. A cada escolha do próximo nó a receber a informação, seleciona-se o nó com o maior valor da métrica de *Betweenness* (ou *Closeness*) entre todos os possíveis.

Para o cálculo da métrica de *Betweenness* (ou *Closeness*), 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 desta métrica pode ser custoso. Como trabalho futuro, pretende-se verificar a possibilidade do cálculo distribuído desta métrica, bem como a aplicação de outras métricas topológicas.

#### 5. Descrição dos Logs

Neste artigo são modelados dois tipos diferentes de sistemas reais. O primeiro reúne informações sobre a mobilidade de usuários entre diferentes pontos de acesso de uma rede *wireless* de um *campus* universitário. O segundo representa as trocas de informações entre os diversos usuários de uma aplicação P2P de difusão de vídeo ao vivo. A seguir, são detalhados cada um destes *logs*.

##### 5.1. Dartmouth

O *log* Dartmouth inclui informações do protocolo SNMP 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.

Para a análise considerada neste artigo, 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 da troca de informações entre os dispositivos. Para a construção do grafo  $\mathcal{G}_t$ , a escolha para o valor  $\Delta$  foi de 3.600 segundos (em linha com resultados mostrados em [Balachandran et al. 2002]), com um total de 8.769 *snapshots* diferentes a serem analisados.

## 5.2. SopCast

O SopCast [Sopcast 2012] 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, i.e, 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 artigo 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  $\Delta = 1s$ , 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 3.601 *snapshots* de 1s, com o total de 334 nós no grafo que modela o sistema. Uma aresta é estabelecida entre dois nós (clientes) da rede sobreposta, se existe, 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 em estudo é um grafo direcionado.

## 6. Resultados

O principal foco dos resultados apresentados neste artigo é avaliar o impacto da dinâmica inerente aos sistemas descritos na Seção 5, bem como o desempenho dos diferentes modelos de difusão apresentados nas Seções 3.4 e 4.

Seja o total de nós alcançados a partir do nó  $j$  denotado por  $TN_j$  e  $|\mathcal{V}_{T_i}|$  o total de nós presentes no grafo no instante de tempo  $T_i$ . Em todos os resultados relativos aos modelos de difusão de informação considera-se a medida de interesse *percentual de nós alcançados em  $T_i$*  ( $\pi_{T_i}$ ), definida por:

$$\pi_{T_i} = \frac{\sum_{\forall j \in \mathcal{V}_{T_i}} TN_j}{|\mathcal{V}_{T_i}| * (|\mathcal{V}_{T_i}| - 1)}.$$

Esta métrica calcula a média dos nós alcançados a cada *snapshot*, considerando cada um dos nós em  $\mathcal{V}_{T_i}$  como nó inicial de difusão. Como busca-se 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 modelo de difusão, opta-se pela média calculada considerando todos os nós como sendo nós iniciais da difusão.

## 6.1. Influência dos Diferentes Conceitos de Dinâmica no desempenho dos modelos SoA e CuT

### 6.1.1. Mobilidade dos nós

Como primeiro conjunto de resultados, é apresentado o impacto da mobilidade dos nós na difusão de informação para a implementação *Epidêmico* dos modelos SoA e CuT. A mobilidade das entidades de um sistema real pode ser considerada como um tipo particular de dinâmica do sistema, sendo esta relacionada com o comportamento das entidades pertencentes ao mesmo. Em termos de representação matemática, esta definição implica na inserção, permanência e retirada de arestas do grafo.

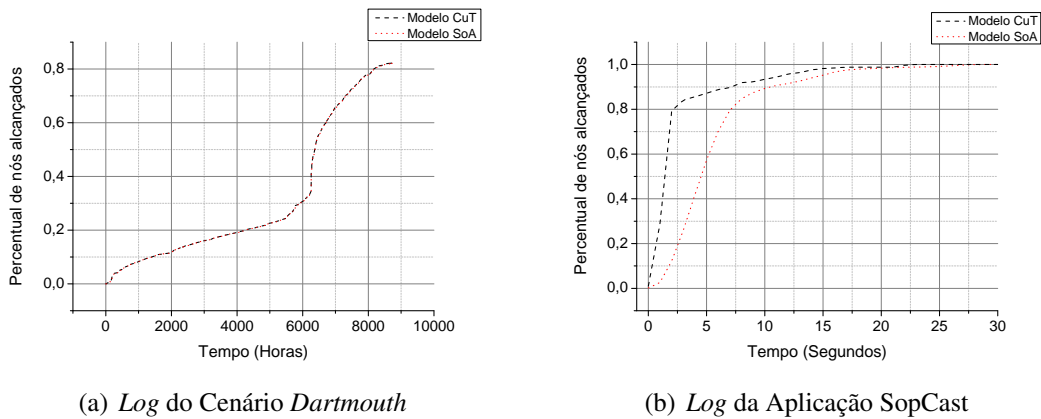
Para o *log Dartmouth* a mobilidade será definida como mudança de localização de um dispositivo entre os diferentes pontos de acessos que formam a rede *wireless*. Para o *log* da aplicação SopCast, é considerado como mobilidade a criação de novas parcerias para troca de porções de vídeo.

A Figura 2 apresenta o resultado da difusão da informação entre as entidades destes dois sistemas reais. Algumas considerações podem ser feitas. Primeiro, o tempo necessário para atingir um percentual significativo de nós ( $> 20\%$ ) do cenário *Dartmouth* é muitas ordens de grandeza maior que a duração de um *snapshot* ( $\Delta = 1h$ ). 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 a dinâmica 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). A segunda consideração está relacionada ao resultado semelhante entre os modelos de difusão 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. Desta forma, os modelos SoA e CuT se comportam de forma análoga, atingindo todos os nós dentro de uma componente ligada.

### 6.1.2. Agregado versus Snapshots

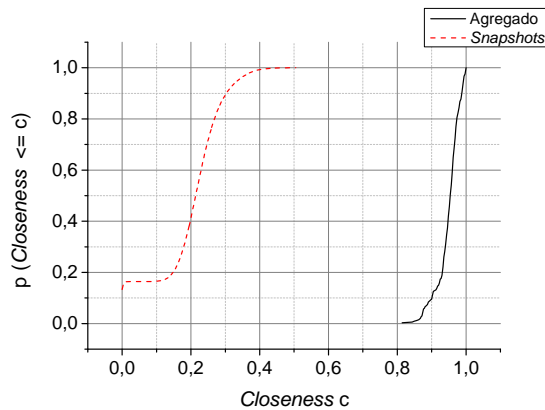
Nesta seção será apresentado um conjunto de resultados para o *log* que representa a aplicação SopCast. Futuramente, será analisado o impacto da utilização da visão em uma única fotografia no sistema real representado no *log Dartmouth*.

Inicialmente, considera-se o cálculo das métricas topológicas realizadas no grafo estático  $G$  e no conjunto de grafos dinâmicos representados por  $\mathcal{G}_T$ . A Figura 3 apresenta a função de distribuição acumulada (CDF) dos valores de *closeness* dos participantes de ambos os cenários. Apresenta-se somente a métrica de *closeness*, visto que para as métricas de grau e *betweenness*, o comportamento é similar. Conforme pode-se verificar, os valores para métrica *closeness* são superestimados. Considerando a análise dinâmica, a



**Figura 2. Dinâmica considerando Mobilidade: Desempenho dos Modelos SoA e CuT (*Epidêmico*).**

métrica possui valores até 0.4. Considerando o grafo agregado  $G$ , a métrica possui valores entre 0.8 e 1.0, com o valor mínimo duas vezes maior que o valor máximo encontrado aplicando a análise dinâmica. Este fenômeno ocorre pois, ao se considerar o grafo agregado, um número maior de caminhos são estabelecidos instantaneamente, influenciando nas métricas relacionadas a caminhos mínimos encontrados entre dois nós.

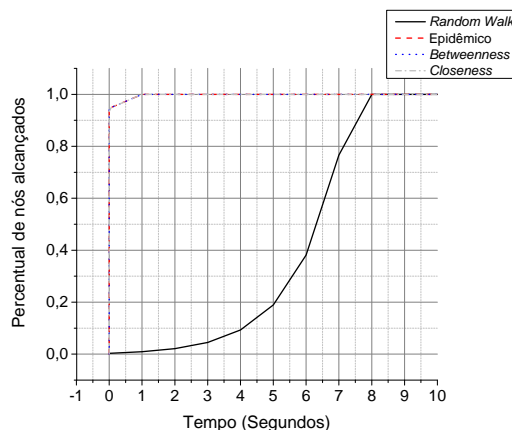


**Figura 3. Métrica de *closeness* - Agregado versus Snapshots.**

A seguir, foca-se no estudo do impacto no desempenho do modelo SoA ao se considerar a representação do sistema em uma única fotografia, através do grafo estático  $G$  e ao se considerar o sistema dinâmico, formada pelos grafos  $\mathcal{G}_t$ . Sejam as implementações *Epidêmico*, RW e *Betweenness Walk*. Vale ressaltar que, para o grafo estático  $G$ , agrupa-se informação de 1h de observação do sistema. Em contrapartida, considerar vários *snapshots* proporciona observação mais fluída das mudanças ocasionadas pela dinâmica da aplicação SopCast. Neste caso, o grafo  $\mathcal{G}_t$  acumula informações de apenas 1s de observação.

A Figura 4 mostra os resultados para a medida de interesse  $\pi_{T_i}$ . No grafo agregado  $G$  todas as arestas que ocorreram em algum instante de tempo dentro do período de observação do sistema são acrescentadas instantaneamente em  $G$ . Assim, tanto a versão *Epidêmico*, RW, *Betweenness Walk* e *Closeness Walk* resultam em um tempo *virtualmente*

menor para a difusão da informação entre os clientes da aplicação SopCast. Nesse caso, a análise dos modelos de difusão no grafo estático resulta em uma visão otimista (e por vezes errada) do desempenho dos processos de difusão.



**Figura 4. Dinâmica considerando Agregado - Modelos de Difusão**

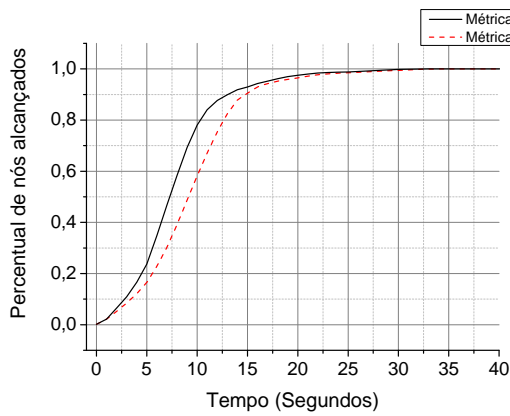
Completando a importância da dinâmica na modelagem, considera-se o caso em que é selecionado um novo nó que irá receber a informação baseado na métrica de *Betweenness* ou *Closeness* calculada somente no grafo estático  $G$ . Em outras palavras, o cálculo é realizado unicamente considerando o grafo agregado e o mesmo valor é utilizado a cada intervalo de tempo  $t$  de difusão da informação ( $T_1 \leq t \leq T_N$ ).

A Figura 5 mostra os resultados para as versões SoA e CuT dos modelos *Betweenness Walk* e *Closeness Walk*. Conforme esperado, ao se aplicar a versão SoA, o processo de difusão é até 20% mais rápido quando recalculamos a métrica a cada novo *snapshot* analisado. No caso da versão CuT, conforme mostrado na Figura 4, os resultados dos modelos de difusão considerando as métricas topológicas se comportam de maneira análoga ao melhor caso, que é a versão *Epidêmico* deste modelo.

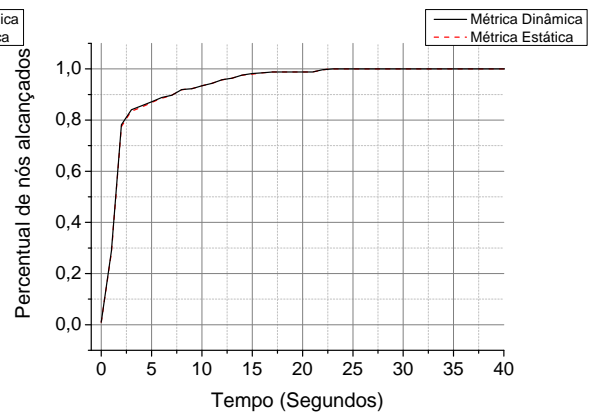
## 6.2. Acelerando o Processo de Difusão

Finalizando o conjunto de resultados, é analisada a importância de incluir, na definição de novos modelos de difusão, características dinâmicas da topologia do grafo que modela o sistema real. A Figura 6 apresenta os resultados da difusão de informação, para a aplicação SopCast, considerando os modelos SoA e CuT nas implementações *Epidêmico*, *Random Walk*, *Betweenness Walk* e *Closeness Walk*. Vale ressaltar que as diferentes implementações serão aplicadas na coleção de grafos  $G_t$  que fornecem a análise dinâmica da aplicação.

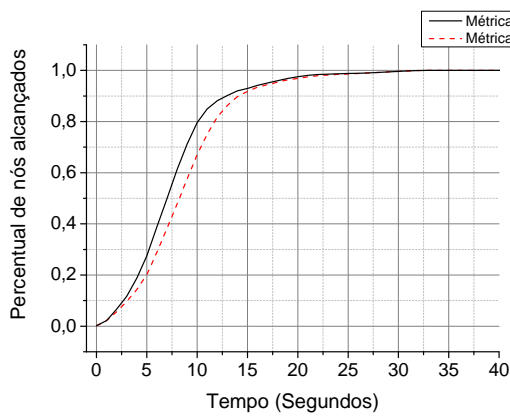
Inicialmente, serão considerados os resultados referentes ao modelo *Store-or-Advance* apresentados no gráfico 6-(a). Conforme esperado, o melhor desempenho é alcançado com a implementação *Epidêmico*, que abrange todos os vizinhos a cada iteração do modelo. A consideração mais interessante refere-se à diferença do tempo de difusão entre o *Random Walk*, *Betweenness Walk* e *Closeness Walk*: ao se incluir o conhecimento da proporção dos caminhos mínimos que passam por um nó e privilegiando os nós com maior valor desta métrica, diminui-se em aproximadamente 20% o tempo para



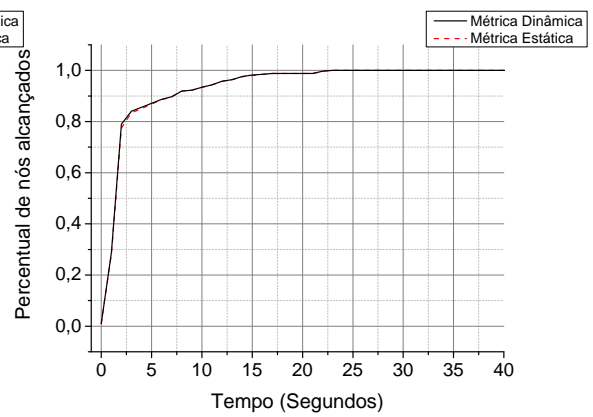
(a) *Betweenness Walk - SoA.*



(b) *Betweenness Walk - CuT.*

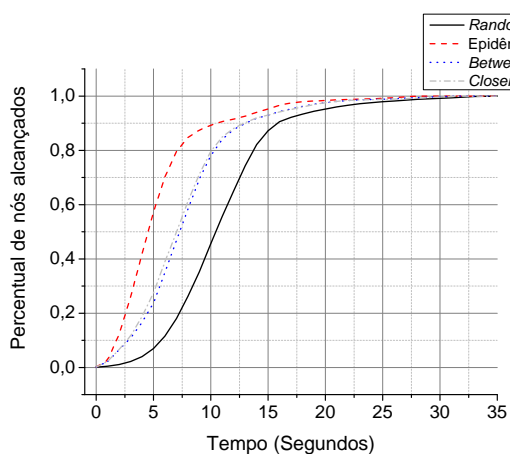


(c) *Closeness Walk - SoA.*

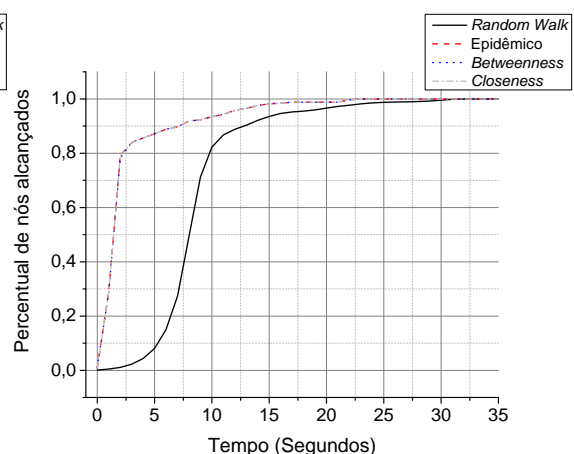


(d) *Closeness Walk - CuT.*

**Figura 5. Impacto da utilização das métricas do agregado em cada *Snapshot* versus o cálculo das métricas do conjunto de grafos  $\mathcal{G}_t$ , com  $T_1 \leq t \leq T_N$ .**



(a) Modelo *Store-or-Advance*



(b) Modelo *Cut-Through*

**Figura 6. Acelerando o Processo de Difusão.**

difundir a informação a 50% dos nós que participam da aplicação SopCast, tanto em relação os modelos *Betweenness Walk* quanto *Closeness Walk*. A pequena diferença entre

os modelos que consideram as métricas topológicas pode ser explicada pela característica desta rede em particular: o diâmetro<sup>1</sup> desta rede é pequeno, aproximadamente igual a 4. Assim pode-se dizer que a maioria dos nós está próxima aos demais em pequenos saltos e que os nós entre si fazem parte da maioria dos caminhos mínimos, resultando em um pequeno ganho do modelo *Closeness Walk*. O ganho do modelo *Closeness Walk* foi de 3.98% em relação ao modelo *Betweenness*.

Em contrapartida, os resultados encontrados na Figura 4 não levam em consideração a dinâmica do sistema. Neste caso, cerca de 95% dos nós já são alcançados no instante inicial de observação. Ao se considerar a dinâmica, para atingir os mesmos 95% dos participantes são necessários 10s. Estes valores são indícios que ignorar a dinâmica do sistema pode levar a uma estimativa errada do valor do tempo de difusão do sistema avaliado.

Para finalizar, considera-se os resultados referentes ao modelo *Cut-Through*, apresentados pela Figura 6-(b). Ao se considerar as métricas *Betweenness* e *Closeness* no processo de difusão, o desempenho se aproxima ao desempenho da implementação *Epidêmico*. Alguns pontos que explicam estes resultados são: primeiro, o CuT difunde a informação a *todos* os nós que pertencem a componente ligada, inclusive aquela que atinge o maior número de nós em um menor número de saltos. Segundo, 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 semelhante, dada a característica do pequeno diâmetro da rede.

## 7. Conclusões e Trabalhos Futuros

Neste artigo são analisados os impactos causados pela dinâmica em redes de comunicação. Em particular, são comparados os modelos que representam o sistema real em estudo com e sem dinâmica. Para tal análise, foi utilizado o problema de como difundir informação nestas redes. Os principais pontos discutidos neste artigo podem ser divididos em dois grupos.

Primeiro o conceito de dinâmica é amplo e claramente influencia o desempenho dos algoritmos de difusão de informação. Esta afirmação é baseada nos resultados onde são aplicados os algoritmos SoA e CuT nos dois cenários de aplicações analisados. Devido ao comportamento pouco dinâmico dos usuários do *log Dartmouth* e de sua estrutura de topológica particular, os dois modelos alcançam desempenho sem diferenças significativas. Em contrapartida, no caso da aplicação *SopCast*, devido a descoberta acelerada de novos nós com os quais se efetuam trocas de vídeo, a exploração das métricas de centralidade dos usuários, permite uma difusão de informação cerca de 80% mais rápida.

Em segundo lugar, a inclusão das características dos nós que formam a rede, aumentam o desempenho dos modelos de difusão. Mais ainda, os ganhos obtidos no tempo de difusão da informação se alteram ao se considerar as mudanças destas métricas ao longo do tempo. Conforme observado, é possível diminuir em 20% o tempo para difundir a informação a uma grande parcela dos destinos quando considera-se as métricas topológicas de *Betweenness* e *Closeness*.

Os resultados apresentados neste artigo são encorajadores e contribuem para di-

---

<sup>1</sup>diâmetro é definido como o maior entre todos os caminhos mínimos

versos questionamentos interessantes, tais como: (i) é possível definir alguma heurística que considera a dinâmica e as características topológicas do sistema real com o objetivo de obter o valor mais próximo possível dos modelos SoA e CuT *Epidêmico*? (ii) como definir o tamanho do intervalo do *snapshot* e inserir o grau máximo de dinâmica na modelagem de redes de comunicação? (iii) qual a melhor métrica topológica a ser considerada, baseando-se no sistema real em estudo? (iv) qual é a melhor definição de dinâmica para um dado sistema real?

Finalizando, como o comportamento dos modelos de difusão estão estreitamente ligados à estrutura topológica da rede, em trabalhos futuros pretende-se abranger diferentes tipos de redes, como redes sociais e de outros tipos de redes de comunicação.

## Referências

- Balachandran, A., Voelker, G., Bahl, P., and Rangan, P. (2002). Characterizing user behavior and network performance in a public wireless lan. In *ACM SIGMETRICS Performance Evaluation Review*, volume 30, pages 195–205. ACM.
- Barbera, M., Stefa, J., Viana, A., de Amorim, M., and Boc, M. (2011). 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*, pages 1–8.
- Basu, P., Bar-Noy, A., Ramanathan, R., and Johnson, M. P. (2010). Modeling and Analysis of Time-Varying Graphs. Published on arXiv.org;cs;arXiv:1012.0260.
- Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R., and Scott, J. (2007). Impact of Human Mobility on Opportunistic Forwarding Algorithms. *IEEE Trans. Mob. Comp*, 6:606–620.
- Hui, P., Crowcroft, J., and Yoneki, E. (2008). Bubble rap: social-based forwarding in delay tolerant networks. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing, MobiHoc '08*, New York, NY, USA. ACM.
- Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H., and Makse, H. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893.
- Lovasz, L. (1993). Combinatorics, Paul Erdos is Eighty. *Bolyai Society Mathematical Studies*, 2:1–46.
- Sopcast (2012). <http://www.sopcast.org/>.
- Tang, S., Lu, Y., Hernández, J. M., Kuipers, F., and Mieghem, P. (2009). Topology dynamics in a p2ptv network. In *Proceedings of the 8th International IFIP-TC 6 Networking Conference, NETWORKING '09*, pages 326–337, Berlin, Heidelberg. Springer-Verlag.
- Wehmuth, K. and Ziviani, A. (2011). Um Novo Algoritmo Distribuído para Avaliação e Localização de Centralidade de Rede. In *Proceedings of X Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPerformance)*.