

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Programa de Pós-Graduação em Ciência da Computação

Eduardo Chinelate Costa

Centralidade de Tempo em Grafos Variantes no Tempo

Juiz de Fora

2015

Eduardo Chinelate Costa

Centralidade de Tempo em Grafos Variantes no Tempo

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora, na área de concentração em Redes de Computadores, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: Alex Borges Vieira

Coorientadora: Ana Paula Couto da Silva

Juiz de Fora

2015

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Costa, Eduardo Chinelate.

Centralidade de Tempo em Grafos Variantes no Tempo / Eduardo Chinelate Costa. – 2015.

70 f. : il.

Orientador: Alex Borges Vieira

Coorientadora: Ana Paula Couto da Silva

Dissertação (Mestrado) – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2015.

1. Grafos Variantes no Tempo. 2. Processo de Difusão de Informação.
3. Centralidade de Tempo. I. Vieira, Alex Borges, orient. II. Silva, Ana Paula Couto da, coorient. III. Título

Eduardo Chinelate Costa

Centralidade de Tempo em Grafos Variantes no Tempo

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora, na área de concentração em Redes de Computadores, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em: 23 de Fevereiro de 2015

BANCA EXAMINADORA

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

Prof. D. Sc. Ana Paula Couto da Silva - Coorientadora
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

AGRADECIMENTOS

Agradeço a Deus, por todos os momentos bons que vivi, pelas pessoas que conheci e por me mostrar que as coisas sempre podem dar certo quando acreditamos nelas.

Agradeço aos meus pais, Bernadete e Emanuel, que são a razão da minha vida. Obrigado por tudo o que vocês me ensinaram e por todo o esforço que vocês tiveram em criar meu irmão e eu, apesar das inúmeras dificuldades impostas pela vida. A vocês, minha eterna gratidão e todo meu amor.

Agradeço a minha namorada Aline por ter sido a melhor companheira que um mestrando poderia ter. Obrigado pela compreensão nos momentos de ausência, pela paciência nos momentos de dificuldade e por todo amor que você me confiou. Sem você, minha vida não teria sentido. Nós ainda temos um longo caminho pela frente, *but you'll never see the end of the road while you're travelling with me*. Eu te amo!

Agradeço ao futuro mestre Leonardo por ser o melhor irmão e amigo do mundo, definitivamente. Obrigado pelo apoio psicológico e intelectual, que somente uma pessoa com um coração de ouro como o seu poderia fazer. Nunca desanime, por mais que as coisas pareçam difíceis. Tenho certeza que vamos conquistar o mundo. Agradeço também a minha cunhada Cristiane, que dividiu conosco, em família, todos os percalços do mundo acadêmico da computação. Amo vocês!

Dedico este trabalho inteiramente ao meu orientador Alex. Não tenho palavras para agradecer por todo esse tempo de trabalho juntos, desde a graduação. Obrigado pela paciência, pela dedicação e pela confiança depositada nesse pesquisador em formação. Jamais esquecerei da sua ajuda e compreensão dos momentos de dificuldade, profissionais e pessoais. Desejo que sua vida aqui seja repleta de luz. Agradeço também a minha primeira orientadora Ana Paula, por sua contribuição na minha formação como aluno e pesquisador, desde o início. Meu carinho a vocês!

Agradeço também ao professor Artur Ziviani e ao Klaus, do LNCC, pelo grande apoio dedicado a mim e a este trabalho. Obrigado por terem me acolhido em seu grupo de pesquisa, quando eu era apenas um recém-formado sem rumo na vida. Sem vocês, nada disso teria acontecido. A vocês, minha gratidão.

Agradeço a todos os meus amigos e a minha grande família. Não posso colocar os nomes de todos aqui, mas lembro de cada um de vocês que contribuíram para que eu me tornasse a pessoa que sou hoje. Agradeço em especial aos companheiros de laboratório.

Agradeço à FAPEMIG pelo suporte financeiro durante os anos de mestrado, sem o qual meu trabalho seria impossível. Agradeço a UFJF por todos esses anos de parceria, e espero que ainda tenhamos uma longa história em comum. Agradeço ao PGCC pela oportunidade, pela estrutura e pela confiança depositada no meu potencial.

“And I wake up blind
Like my dreams were too bright
And I lost my regard
For the good things that I had”
Pour Le Monde (Crowded House)

RESUMO

Atualmente, há um grande interesse em investigar a dinâmica em Grafos Variantes no Tempo (GVTs). Esses grafos contemplam a evolução temporal, tanto de nós, quanto de arestas. Nesse cenário, de maneira similar a grafos estáticos, o conceito de centralidade geralmente se refere a métricas que avaliam a importância relativa dos vértices. Entretanto, GVTs possibilitam a avaliação da importância dos instantes de tempo (ou estados) de um grafo ao longo de sua existência. Determinar instantes de tempo importantes nesse contexto pode ter aplicações práticas fortes, sendo particularmente úteis para definir melhores momentos para difusão, gerar modelos e prever o comportamento de GVTs. Neste trabalho, nós definimos *Centralidade de Tempo* em Grafos Variantes no Tempo. A centralidade de tempo avalia a importância relativa dos instantes de tempo. São apresentadas duas métricas de centralidade de tempo voltadas a processos de difusão de informação e uma métrica baseada na disposição das conexões da rede. As métricas foram avaliadas em um conjunto de dados real. Os resultados mostram que os instantes de tempo melhor classificados, de acordo com as métricas criadas, podem tornar o processo de difusão mais rápido e eficiente. Comparado com uma escolha aleatória, o processo de difusão iniciado nos instantes de tempo mais bem classificados pode ser até 2,5 vezes mais rápido, e também pode atingir praticamente o dobro do número de nós na rede em alguns casos.

Palavras-chave: Grafos Variantes no Tempo. Processo de Difusão de Informação. Centralidade de Tempo.

ABSTRACT

Currently, there is a great interest in investigating dynamics in Time-Varying Graphs (TVGs). These graphs contemplate the temporal evolution, both nodes and edges. In this scenario, similar to static graphs, centrality usually refers to metrics that assess the relative importance of vertices. However, in TVGs it is possible to assess the importance of time instants (or states) of a graph throughout its existence. Determining important time instants in this context may have strong practical applications and is particularly useful for defining best times to spread, generate models and predict the behavior of TVGs. In this paper, we define time centrality in Time-Varying Graphs. Time centrality evaluates the relative importance of time instants. We present two time centrality metrics focused on information dissemination processes and another based on layout of network connections.. We evaluate metrics we define relying in a real dataset from an hospital environment. Our results show that the best classified time instants, according to created metrics, can make a faster and more efficient diffusion process. Compared to a random choice, the diffusion process starting at best rated time instants can up to 2.5 times faster, and it also can reach almost double the number of nodes in the network in some cases.

Key-words: Time-Varying Graphs. Information Diffusion Process. Time Centrality.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo baseado em <i>snapshots</i> representado pelo modelo unificado . . .	23
Figura 2 – Conversão de intervalos de tempo contínuos em GVTs discretos	24
Figura 3 – Representação de um GVT baseado em arestas espaciais e temporais .	25
Figura 4 – Representação de um GVT baseado em arestas temporais e mistas . .	26
Figura 5 – Exemplo para a representação de um GVT	29
Figura 6 – Forma matricial do tensor de adjacência do GVT W (WEHMUTH et al., 2014a)	31
Figura 7 – Representação estrutural do GVT W	42
Figura 8 – Comparando diferentes processos de difusão	45
Figura 9 – Representação estrutural do GVT N e difusão iniciada no nó $(0, t_0)$. .	47
Figura 10 – Comparando diferentes processos de difusão	49
Figura 11 – Função de distribuição cumulativa do Grau do Tempo	55
Figura 12 – Tamanho do <i>snapshot</i> : 30 segundos	56
Figura 13 – Tamanho do <i>snapshot</i> : 1 minuto	56
Figura 14 – Tamanho do <i>snapshot</i> : 2 minutos	56
Figura 15 – Tamanho do <i>snapshot</i> : 4 minutos	56
Figura 16 – Função de distribuição cumulativa da Variação do Grau do Tempo . .	57
Figura 17 – Função de distribuição cumulativa do Tempo de Cobertura	58
Figura 18 – Cobertura da rede: 10%	59
Figura 19 – Cobertura da rede: 20%	59
Figura 20 – Cobertura da rede: 30%	59
Figura 21 – Cobertura da rede: 40%	59
Figura 22 – Cobertura da rede: 50%	59
Figura 23 – Cobertura da rede: 60%	59
Figura 24 – Tempo de Cobertura (10% de Cobertura da rede): 10 Melhores × Demais instantes	60
Figura 25 – Tempo de Cobertura: matriz de similaridade entre as classificações . .	61
Figura 26 – Função de distribuição cumulativa da Centralidade de Cobertura . . .	62
Figura 27 – Número fixo de passos: 500	63
Figura 28 – Número fixo de passos: 1000	63
Figura 29 – Número fixo de passos: 1500	63
Figura 30 – Número fixo de passos: 2000	63
Figura 31 – Centralidade de Cobertura (2000 passos): 10 Melhores × Demais instantes	64
Figura 32 – Centralidade de Cobertura: matriz de similaridade entre as classificações	64

LISTA DE ABREVIATURAS E SIGLAS

GVTs	Grafos Variantes no Tempo
RFID	Identificação por radiofrequência
MOSAR	<i>Mastering hOSpital Antimicrobial Resistance and its spread</i>
AMRB	<i>AntiMicrobial-Resistant Bacteria</i>
SIR	<i>Suscetível-Infetado-Removido</i>
SIS	<i>Suscetível-Infetado-Suscetível</i>
RW	<i>Random Walk</i>
CDF	<i>Cumulative Distribution Function</i>
CCDF	<i>Complementary Cumulative Distribution Function</i>
CoRR	<i>Computing Research Repository</i>

LISTA DE SÍMBOLOS

\forall	Para todo
\exists	Existe
\in	Pertence
\notin	Não pertence
\subset	Está contido
\subseteq	Contido em ou igual a
Σ	Somatório
\Rightarrow	Implica em

SUMÁRIO

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	GRAFOS VARIANTES NO TEMPO (GVTs)	14
2.1.1	Definição	14
2.1.2	Formalismo	15
2.1.3	Aplicações	17
2.2	MÉTRICAS TRADICIONAIS DE CENTRALIDADE	19
2.2.1	Grau	19
2.2.2	<i>Betweenness</i>	20
2.2.3	<i>Closeness</i>	21
2.3	MODELOS DE REPRESENTAÇÃO DE GVTs	21
2.3.1	Modelos baseados em <i>snapshots</i>	22
2.3.2	Modelos baseados em intervalos de tempo contínuos	23
2.3.3	Modelos baseados em arestas espaciais e temporais	25
2.3.4	Modelos baseados em arestas temporais e mistas	25
3	MODELO DE REPRESENTAÇÃO UNIFICADO DE GVTs .	27
3.1	REPRESENTAÇÃO ESTRUTURAL DE UM GVT	29
3.1.1	Tensor de Adjacência de um GVT	29
3.1.2	Representação Matricial do Tensor de Adjacência de um GVT	30
4	TRABALHOS RELACIONADOS	33
4.1	MODELOS DE DIFUSÃO DE INFORMAÇÃO EM REDES	36
4.1.1	Modelos Epidêmicos	36
4.1.1.1	<i>Modelo Epidêmico SIR</i>	37
4.1.1.2	<i>Modelo Epidêmico SIS</i>	38
4.1.2	Modelo <i>Random Walk</i>	38
5	CENTRALIDADE DE TEMPO	40
5.1	MÉTRICAS DESENVOLVIDAS PARA CENTRALIDADE DE TEMPO	40
5.1.1	Grau do Tempo	41
5.1.2	Tempo de Cobertura	44
5.1.3	Centralidade de Cobertura	47
6	APLICAÇÃO E ANÁLISE DE CENTRALIDADE DE TEMPO	53
6.1	CONJUNTO DE DADOS MOSAR	53

6.2	GRAU DO TEMPO	54
6.3	VARIAÇÃO DO GRAU DO TEMPO	55
6.4	TEMPO DE COBERTURA	56
6.5	CENTRALIDADE DE COBERTURA	61
7	CONCLUSÃO E TRABALHOS FUTUROS	65
	REFERÊNCIAS	67

1 INTRODUÇÃO

O conceito de centralidade na Teoria de Grafos e na análise de redes geralmente se refere às métricas que avaliam a importância relativa de vértices (nós) em um grafo (rede). Diferentes modos de medir a centralidade de nó foram propostos por décadas, sendo realizada por diferentes pontos de vista (FREEMAN, 1978, 1977; SABIDUSSI, 1966). Exemplos incluem o uso de centralidade de nó para avaliar a robustez da rede frente à fragmentação ou para identificar os nós mais importantes para difusão eficiente de informação (BUI-XUAN et al., 2003; KHELIL et al., 2002).

Mais recentemente, existe um interesse cada vez maior na investigação da dinâmica de redes, ou seja, quando a estrutura da rede (nós e arestas) pode variar ao longo do tempo. Isto está relacionado ao estudo de Grafos Variantes no Tempo (GVTs) (NICOSIA et al., 2013; KIM and ANDERSON, 2012; TANG et al., 2010a). Consequentemente, vários esforços recentes investigam novas definições de centralidade para capturar a importância relativa de nós no contexto de um GVT, como a identificação dos nós mais estratégicos para o processo de difusão de informação.

Para algumas aplicações modeladas em cenários de GVT, mais importante do que identificar os nós centrais sob uma determinada definição, é identificar os principais instantes de tempo para a tomada de certas decisões. Isto pode ser de extrema importância para obter resultados mais efetivos. Por exemplo, é mais interessante determinar o melhor instante de tempo no qual uma vacina pode ser aplicada em uma população para atingir melhores respostas contra uma epidemia.

Nesse sentido, nós definimos a *Centralidade de Tempo* em um GVT. De maneira similar à centralidade de nó, a centralidade de tempo avalia a importância de um dado instante de tempo em um GVT e pode ser determinada de diferentes maneiras. Em um processo de difusão de informação, poderia ser mais estratégico identificar o melhor momento para iniciar uma difusão mais rápida do que identificar o melhor nó para iniciar uma difusão.

Até onde sabemos, este é o primeiro trabalho que aborda a centralidade de tempo em GVTs. De fato, os trabalhos anteriores estão mais interessados em qualificar a importância dos nós em uma rede do que em determinar os instantes de tempo mais importantes no contexto de um GVT.

Para identificar os instantes de tempo importantes, é necessário que haja previamente algum conhecimento sobre o estado da rede e sua evolução ao longo do tempo. Isto não é uma limitação, visto que muitos eventos modelados por GVTs são notoriamente cíclicos, como mobilidade em redes de transportes e surtos epidêmicos de doenças. Consequentemente, a Centralidade de Tempo seria útil na construção de modelos de previsão. Com base em propriedades de instantes de tempo importantes no passado, uma vez que

tais propriedades tornassem a ocorrer, o modelo de previsão em questão seria capaz de indicar instantes de tempo importantes em potencial.

A fim de abordar a centralidade de tempo em GVTs, são apresentadas três métricas: *Grau do Tempo*, *Tempo de Cobertura* e *Centralidade de Cobertura*. A primeira é baseada na disposição das conexões da rede e as últimas são voltadas a processos de difusão de informação. As métricas desenvolvidas foram avaliadas em um conjunto de dados real, obtido de uma rede de contatos presenciais em um ambiente hospitalar.

Os resultados mostram que os instantes de tempo melhor avaliados, de acordo com as métricas desenvolvidas, podem realizar um processo de difusão mais rápido e eficiente. Por exemplo, quando comparado a uma escolha aleatória, um processo de difusão pode atingir 20% da rede até 2,5 vezes mais rápido. Já a cobertura da rede em 1500 passos do processo de difusão praticamente dobra com relação aos instantes de tempo escolhidos aleatoriamente.

A dissertação está organizada da seguinte forma: no capítulo 2, são mostrados os conceitos fundamentais de Grafos Variantes no Tempo, suas diferenças com relação a redes estáticas, definições e áreas de aplicação; além da definição das principais métricas tradicionais de centralidade. No capítulo 3, são descritos o modelo de representação de GVTs utilizado neste trabalho e sua formalização. O capítulo 4 apresenta um conjunto de trabalhos encontrados na literatura referente a métricas de centralidade temporal em redes e modelos de difusão de informação. No capítulo 5, é definida a *Centralidade de Tempo* em Grafos Variantes no Tempo e são formalizadas as três métricas de avaliação desenvolvidas. No capítulo 6, são mostrados os resultados das métricas de avaliação frente ao conjunto de dados. No capítulo 7, a dissertação é concluída e são relatados possíveis trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais de Grafos Variantes no Tempo (GVTs) e centralidade de nó. Na seção 2.1, são discutidas as diferenças entre GVTs e rede estáticas e as diferentes definições formais e aplicações de GVTs encontradas na literatura. Além disso, são descritas brevemente muitas das áreas de aplicação de GVTs e estudos relacionados ao tema. Na seção 2.2, são apresentadas as principais métricas de centralidade de nó para grafos estáticos. Além disso, na seção 2.3, são descritas as diferentes classes de modelos de representação de GVTs.

2.1 GRAFOS VARIANTES NO TEMPO (GVTs)

2.1.1 Definição

Na literatura, os conceitos de grafo e rede estão geralmente relacionados. De maneira mais direta, um grafo é tido como uma representação matemática de uma rede. De acordo com EVANS and MINIEKA (1992), uma rede é um grafo com um ou mais números associados por arestas ou arcos. Segundo NEWMAN (2008), na literatura matemática, uma rede, também chamada de grafo, é composta por pontos, conhecidos como nós ou vértices, e linhas que os conectam, geralmente chamadas de arestas.

Mais formalmente, um grafo G pode ser definido pelo par de conjuntos (V, E) , onde V é um conjunto finito e não vazio de vértices (ou nós) e E é um conjunto de pares não ordenados de V , chamados arestas (ou conjunto de pares ordenados de V , chamados arcos). Cada aresta $e \in E$ é denotada pelo par de vértices $e = (u, v)$, onde $u, v \in V$.

Entretanto, muitos dos sistemas complexos reais encontrados em diferentes campos da ciência, tais como ciências sociais, biologia, computação, economia e telecomunicações, podem ser representados por meio de redes e são notoriamente dinâmicos. Sendo assim, a definição estática de grafos pode não descrever adequadamente o comportamento desses sistemas, já que não consideram o tempo como um fator fundamental na relação entre as entidades participantes da rede. E, de acordo com NICOSIA et al. (2013), o tempo deve ser considerado como outra dimensão (aspecto) de um sistema, de tal forma que ele deve ser incluído na definição formal de um grafo.

De acordo com WEHMUTH et al. (2014a), Grafos Variantes no Tempo (GVTs) são grafos nos quais nós ou arestas podem variar no tempo. NICOSIA et al. (2011) definem Grafos Variantes no Tempo como um modelo natural para sistemas em rede, em que as relações entre os nós são intrinsecamente dinâmicas e variam ao longo do tempo, onde as conexões aparecem e desaparecem em pontos específicos no tempo e são, muitas vezes, recorrentes. CASTEIGTS et al. (2011) consideram GVTs como um conjunto V de entidades (nós), um conjunto E de relações entre tais entidades (arestas) que ocorrem

ao longo de um intervalo de tempo T , conhecido como tempo de vida útil do sistema. No trabalho de TANG et al. (2010b), GVTs são descritos como uma sequência discreta e ordenada de T grafos estáticos (orientados ou não), onde T é o tamanho da sequência e cada grafo representa a estrutura de uma rede dinâmica em um determinado instante de tempo. Conforme NICOSIA et al. (2013), GVTs são, em termos matemáticos, sistemas que consistem em um conjunto E de ligações estabelecidas entre elementos de um conjunto V de nós durante um intervalo de observação $[0, T]$.

Além disso, é importante notar que, por meio do ajuste do tamanho da janela de tempo usada para construir cada instante de tempo (definição de intervalo entre instantes de tempo), é possível obter diferentes representações de um mesmo sistema em diferentes escalas temporais (NICOSIA et al., 2013). Seja o intervalo entre os instantes de tempo de um GVT dado por Δt . No caso limite, quando Δt tende a zero 0, obtém-se uma sequência infinita de grafos, onde cada um dos grafos corresponde à configuração de contatos em um dado instante de tempo t . Esta sequência de grafos pode incluir um certo número de grafos vazios, correspondentes a períodos em que não há conexões registradas. Em contrapartida, se $\Delta t = T$, o GVT se degenera em um grafo agregado não ponderado correspondente, onde toda a informação temporal é perdida.

Resumidamente, o conceito de GVT se refere a grafos nos quais as relações entre entidades de um conjunto de nós e entidades de um conjunto de arestas (orientadas ou não) são descritas de acordo com um aspecto temporal - contínuo ou discretizado em intervalos bem definidos - de maneira sequencial, o que sugere uma noção de dinamicidade.

2.1.2 Formalismo

Existem, na literatura, algumas definições formais para GVTs sob diferentes perspectivas. Segundo TANG et al. (2010b), um sistema pode ser descrito em seu tempo de amostragem máximo como um grafo variante no tempo, ou seja, uma sequência discreta (um conjunto ordenado) G_1, G_2, \dots, G_T de T grafos não orientados ou orientados, onde $|T|$ é o comprimento da sequência. A rede tem N nós, de tal forma que as ligações entre eles podem variar ao longo do tempo. Portanto, o GVT pode ser representado por meio de uma matriz de adjacência dependente do tempo $A(t)$, com $t = 1, \dots, T$, e dimensões $N \times N$, onde $a_{ij}(t)$ são os elementos da matriz de adjacência do t -ésimo grafo.

Conforme HOLME and SARAMAKI (2012), uma rede temporal é dividida em duas classes. Na primeira, há um conjunto de N nós $\in V$ interagindo uns com os outros em determinados instantes de tempo, e as durações das interações são negligenciadas. Neste caso, o sistema pode ser representado por uma sequência de contatos - um conjunto de contatos C , como uma tupla (i, j, t) , onde $i, j \in V$ e t indica o instante de tempo. De forma equivalente, pode-se representar o sistema por um conjunto de arestas E , e para $e \in E$, um conjunto não vazio de tempos de contatos e $T = t_1, \dots, t_n$. Na segunda classe,

discute-se os intervalos entre grafos, ou seja, as arestas que não estão ativas em um instante de tempo específico, mas sim ao longo de um conjunto de intervalos de instantes de tempo $Te = (t_1, t'_1), \dots, (t_n, t'_n)$, onde os elementos do conjunto indicam os períodos de atividade dos intervalos.

Nos trabalhos de CASTEIGTS et al. (2011); SANTORO et al. (2011), um GVT é representado por um conjunto V de entidades (nós), um conjunto E de relações entre tais entidades (arestas) e um alfabeto L que representa qualquer propriedade que uma relação tal possa ter (rótulo); isto é, $E \subseteq V \times V \times L$. A definição de L é de domínio específico, e, conseqüentemente, é deixada em aberto - um rótulo pode representar, por exemplo, a intensidade da relação em uma rede social, um tipo de veículo numa rede de transporte, ou um meio em particular numa rede de comunicação; em alguns contextos, L pode estar vazio (e, assim, possivelmente omitido). Assume-se que L pode conter elementos multivalorados (por exemplo: largura de banda, tipo de enlace, disponibilidade de criptografia, entre outros).

As relações entre as entidades ocorrem ao longo de um intervalo de tempo $\Delta T \subseteq T$ chamado de vida útil do sistema. O domínio temporal T é geralmente assumido como N para sistemas de tempo discreto ou R^+ para sistemas de tempo contínuo. A dinâmica de um sistema pode ser descrita por um GVT $G = (V, E, T, \rho, \gamma)$, onde

- $\rho : E \times T \implies \{0, 1\}$, é chamada função de presença, que indica se uma determinada aresta está disponível em um determinado instante de tempo.
- $\gamma : E \times T \implies T$, é chamada função de latência, que indica o tempo gasto para se percorrer uma determinada aresta, a partir de um determinado instante de tempo.

Ainda segundo CASTEIGTS et al. (2011), uma série de trabalhos de análise em redes dinâmicas simplesmente ignoraram γ , ou assumem um cenário de tempo discreto onde cada instante de tempo implicitamente corresponde a um γ constante.

Todavia, há trabalhos na literatura que apresentam um outro ponto de vista com relação à dinâmica de um sistema representado por um GVT. Como descrito em WEHMUTH et al. (2014a), o modelo de representação de um GVT pode ser dado como um objeto $H = (V, E, T)$, onde V é o conjunto de nós, T é o conjunto finito de instantes de tempo para os quais o GVT é definido, e $E \subseteq V \times T \times V \times T$ é o conjunto de arestas. Uma aresta dinâmica e em um GVT H é definida como uma quadrupla ordenada $e = (u, t_a, v, t_b)$, onde $u, v \in V(H)$ são, respectivamente, nós de origem e destino, enquanto que $t_a, t_b \in T(H)$ são os instantes de tempo de origem e de destino.

Observe que esta definição de arestas dinâmicas aprimora, de certa forma, as funções de presença e de latência caracterizadas no modelo anterior. Apenas com os parâmetros

temporais t_a e t_b é possível determinar e quantificar o intervalo entre os instantes de tempo nos quais a aresta está inserida e, conseqüentemente, a sua latência.

2.1.3 Aplicações

Grafos e redes complexas têm sido usados para estudar diversos fenômenos humanos e naturais (KOSTAKOS, 2009). Relações humanas mudam com o tempo, uma vez que os indivíduos perdem contato com antigos conhecidos, estabelecem novas parcerias, ou movem-se sobre o espaço geográfico; comunicação em redes feitas pelo homem, tais como conexões de máquinas e interações sociais através da Internet, ocorrem em pontos específicos no tempo; novos enlaces são estabelecidos na *World Wide Web*, enquanto outros desaparecem (TANG et al., 2010b). Muitos desses fenômenos podem ser representados por meio de GVTs. HOLME and SARAMAKI (2012); SANTORO et al. (2011); CASTEIGTS et al. (2011) descrevem diversas áreas de aplicação de GVTs e estudos relacionados:

- *Comunicação ponto a ponto.* Os registros de comunicação eletrônica de um-para-um são particularmente adequados para a abordagem de rede temporais, especialmente no contexto da dinâmica de espalhamento de informações ou vírus eletrônicos. Tais dados vêm frequentemente ou sob a forma de listas de mensagens a partir de uma pessoa para outra, num ponto no tempo, ou de um diálogo entre duas pessoas dentro de um intervalo de tempo. O primeiro tipo contém redes de mensagens de e-mail, mensagens de texto de telefone celular, mensagens instantâneas e mensagens em fóruns online. Telefonemas não são instantâneos, mas tem uma duração específica, e assim pode ser considerado do segundo tipo.
- *Disseminação Broadcast.* A comunicação *broadcast*, em contraste com a comunicação ponto-a-ponto, é um outro tipo de espalhamento de informações entre os seres humanos que poderia se beneficiar de uma abordagem de rede temporal. Nesse caso, a comunicação pode ser feita por meio de redes sociais, blogs e microblogs (como *Twitter* ou *Facebook*). YASSERI et al. (2011) levam em consideração a dimensão de tempo de uma forma interessante em sua análise dos padrões de atividade editorial na *Wikipedia*: esses padrões de atividade podem ser utilizados para estimar a distribuição geográfica dos editores.
- *Proximidade física.* Padrões de proximidade dos seres humanos - dados sobre quem está perto de quem e em que momento - são importantes tanto para a compreensão da propagação de organismos patogênicos quanto para divulgação de informação boca-a-boca (*sic*). Dispositivos munidos de identificação por radiofrequência (RFIDs) foram utilizados para a medição de proximidade em redes temporais e dinâmicas de pacientes (ISELLA et al., 2011), escolares (STEHL et al., 2011b) e participantes de conferências (STEHL et al., 2011a).

- *Biologia celular.* Um dos sistemas biológicos que é notavelmente dinâmico é conhecido como interactoma - o conjunto de interações moleculares em uma célula (ITO et al., 2001; LI et al., 2004). Os nós da interactoma são proteínas ou moléculas mais leves que podem se ligar para executar funções biológicas. Frequentemente, estas interações são representadas como um grafo estático. No entanto, grande parte da funcionalidade biológica vem do fato de que as ligações não estão ativas todo o tempo. Outra rede biológica que muda ao longo do tempo é o metabolismo - o conjunto de reações químicas que ocorrem em um organismo saudável (IDEKER et al., 2001). Os nós em redes metabólicas são espécies moleculares que estão ligadas se estiverem envolvidas na mesma reação química. Em um dado momento e em uma dada localização subcelular, apenas uma parte de todo o sistema de reação bioquímica está ativa. Esta situação muda com o tempo, e redes temporais podem potencialmente capturar a sua dinâmica.
- *Computação distribuída.* Uma grande parte dos desenvolvimentos teóricos iniciais em redes temporais vem da ciência da computação. Há muitos tipos diferentes de sistemas de computação distribuída, mas todos consistem de unidades computacionais bastante independentes espalhadas por sobre a rede (GHOSH, 2007). Uma vez que os cálculos são executados em paralelo para difundir as informações entre as unidades, elas normalmente necessitam operar informações com diferentes. Para estudar esse sistema, um problema central é estimar e controlar a idade de uma informação que é acessível para os nós.
- *Redes de transporte.* Um exemplo de sistema modelado por redes temporais é o de transporte aéreo (WANG et al., 2011; CHOI et al., 2006). Os vértices são os aeroportos, as arestas dinâmicas são os voos regulares entre dois aeroportos, determinada pelos tempos de partida e chegada. Esse mesmo esquema se aplica a outros sistemas de transporte, como ônibus, trens e metrô, ou na interligação entre eles (sistemas multicamadas).
- *Redes neurais e cerebrais.* As redes de conexões neurais representam outra classe de redes biológicas que podem se beneficiar da abordagem de redes temporais. Existem vários níveis de conectividade estrutural e temporal, desde padrões de *spiking* (impulsos) de neurônios individuais até conexões fisiológicas ou funcionais menos aperfeiçoadas entre áreas do cérebro. As ligações funcionais refletem as propriedades da conectividade da rede anatômica subjacente (rede de substrato estrutural) entre áreas cerebrais interpostas por feixes de fibras neuronais. Geralmente, em redes funcionais cerebrais temporais, as ligações temporais representam a dinâmica de tempo das ativações simultâneas de áreas cerebrais - enquanto a rede de substrato estrutural é estática em tais escalas de tempo, as ativações de ligações funcionais variam no tempo.

- *Redes ecológicas.* Redes ecológicas capturam as interações entre as espécies ou outras categorias de organismos e, em muitos aspectos, são consideradas dinâmicas. Elas podem mudar, por exemplo, com as estações do ano ou conforme os organismos passem por diferentes fases de seus ciclos de vida (e, portanto, têm diferentes capacidades e necessidades com relação a sua interação com os outros).
- *Redes de Mobilidade Oportunista.* Como dispositivos móveis se tornam cada vez mais equipados com recursos de rádio de curto alcance, é possível explorar as redes criadas por meio de sua mobilidade para usos que são possivelmente externos e alheios aos usuários de tais dispositivos.
- *Redes de sensores sem fio.* As aplicações mais comuns em redes de sensores sem fio são funções de monitoramento e de comunicação. Uma aresta está presente entre dois nós sensores sempre que a distância entre eles for menor que o raio de alcance da transmissão de dados, determinado pelos dispositivos transceptores. A dinâmica da rede é determinada pela movimentação dos sensores no ambiente de interesse, fazendo com que as ligações entre os nós sensores variem ao longo do tempo (LEWIS, 2004; AKYILDIZ and VURAN, 2010).
- *Redes complexas do mundo real.* O problema fundamental nesta área é a definição de modelos matemáticos capazes de capturar e reproduzir propriedades observadas na dinâmica real das redes. KOSTAKOS (2009) apresenta um estudo baseado na dinâmica do comportamento de indivíduos baseado em um grande conjunto de dados sobre trocas de *email*. O remetente e o receptor são os nós, os *emails* trocados entre eles são as arestas e os instantes de tempo são as datas nas quais os *emails* foram enviados. A rede varia a partir do momento em que dois indivíduos não trocam *emails* de maneira regular.

2.2 MÉTRICAS TRADICIONAIS DE CENTRALIDADE

Sucintamente, uma medida de centralidade de nó tenta quantificar a importância relativa de um nó em uma rede, de acordo com alguns aspectos estruturais. Na literatura, diferentes métricas de centralidade foram formalizadas, dentre as quais três são amplamente discutidas: Grau (FREEMAN, 1978), *Betweenness* (FREEMAN, 1977) e *Closeness* (SABIDUSSI, 1966).

2.2.1 Grau

De acordo com FREEMAN (1978), a mais simples e, talvez, a mais intuitiva concepção de centralidade de um nó tem relação com o seu grau. O Grau de um nó é a dado pelo número de ligações que ele estabelece com outros nós em uma rede.

Seja um grafo não orientado $G = (V, A)$, onde V é um conjunto finito de nós que compõem o grafo e A é um conjunto de pares não ordenados de V , os quais são denominados de arestas. O grau de um nó $v \in V$, corresponde ao número de arestas que incidem sobre v .

Em contrapartida, seja considerado um grafo orientado $H = (V, A)$, onde A é um conjunto de pares ordenados de V , que são denominados de arcos. Nesse caso, o conceito de grau é decomposto em grau de saída e grau de entrada. O grau de saída de um nó v equivale ao número de arcos que partem de v e o grau de entrada de um nó v equivale ao número de arcos que chegam a v . O grau total do nó é igual à soma dos graus de saída e entrada, desconsiderando-se geralmente os casos onde há laços (arestas ou arcos que relacionam um nó a ele próprio).

Adicionalmente, BARRAT et al. (2004) estendem a concepção de grau para a soma dos pesos de arestas ou arcos em grafos ponderados em vez de considerar somente o número de arestas ou arcos.

A centralidade de grau de um nó v , propriamente dita, é uma medida relativa ao número de arestas incidentes em v , nos grafos não orientados, ou à soma do número de arcos que chegam e partem de v , em grafos orientados. De acordo com OPSAHL et al. (2010), para um grafo $G = (V, E)$ não orientado, esta medida pode ser formalizada como:

$$d_c(v) = \sum_{u \in V \setminus v} A_{v,u}$$

onde v é o nó; u representa todos os nós do grafo G , exceto o nó v ; N é o número de nós do grafo; A é a de matriz de adjacência de G , na qual a célula $A_{v,u}$ representa o valor do elemento na matriz. O valor é definido como 1, caso haja uma ligação entre v e u , e 0, caso contrário.

Para normalizar a centralidade de grau, a fim de possibilitar a comparação entre redes de tamanhos diferentes, é necessário dividir o grau do nó em questão pelo máximo grau possível que um nó pode assumir, que intuitivamente é igual a $N - 1$. Dessa forma, a centralidade de grau normalizada de um nó v pode ser dada por:

$$d'_c(v) = \frac{d_c(v)}{(N - 1)}$$

2.2.2 *Betweenness*

O conceito de *Betweenness* (ou centralidade de intermediação) foi introduzido por FREEMAN (1977). O *Betweenness* de um nó v é uma medida que retorna a fração dos caminhos mínimos que ligam quaisquer pares de nós de um grafo e que passam pelo nó v (NEWMAN, 2005). Em outras palavras, *Betweenness* calcula basicamente o número de

caminhos geodésicos que passam através de um nó v (BORGATTI, 2005), considerando todos os caminhos mínimos entre quaisquer dois nós de uma rede.

Formalmente, seja $G = (V, A)$ um grafo, onde V é um conjunto finito de nós que compõem o grafo e A é o conjunto de arestas. Considere um par de nós s e $t \in V$ e o nó v . A centralidade de *Betweenness* do nó v é dado por:

$$b_c(v) = \sum_{s \neq t \neq v} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}},$$

onde $\sigma_{s,t}$ é o número caminhos mínimos entre os nós s e t , e $\sigma_{s,t}(v)$ é o número de caminhos mínimos entre os nós s e t , passando pelo nó v .

2.2.3 Closeness

O conceito de *Closeness* (ou centralidade de proximidade) de um nó, introduzido por SABIDUSSI (1966), se refere à soma das menores distâncias entre um nó e todos os demais nós da rede que podem ser alcançados por ele. Em outras palavras, *Closeness* tenta quantificar a proximidade entre um nó específico e todos outros nós da rede.

Dessa forma, conforme mostrado em NEWMAN (2008), a centralidade de *closeness* de um nó v poderia ser dada por:

$$c_c(v) = \sum_{u \in V \setminus v} \frac{1}{dG(v, u)},$$

onde t são todos os nós da rede, exceto o nó v , e dG é a distância geodésica entre os nós v e t .

Entretanto, esta medida reflete um ponto de descentralidade ou centralidade inversa, uma vez que cresce conforme os nós estejam mais afastados, e centralidade, neste contexto, deveria significar proximidade (FREEMAN, 1978). Portanto, uma abordagem mais compatível com tal circunstância seria:

$$c_c(v) = \sum_{u \in V \setminus v} dG(v, u),$$

que garante a proporcionalidade inversa entre distância e proximidade, ou seja, quanto menor a distância entre os nós, maior a centralidade de *Closeness*.

2.3 MODELOS DE REPRESENTAÇÃO DE GVTs

Existem diferentes modelos na literatura que podem representar GVTs. Ao propor seu modelo unificado de representação de GVTs, WEHMUTH et al. (2014a) fazem uma

revisão dos modelos já existentes, agrupados em diferentes classes, a fim de mostrar que as características peculiares de cada um podem ser encontradas conjuntamente no modelo proposto. Segundo os próprios autores, os modelos anteriores não são capazes de representar uns aos outros e não têm o mesmo alcance de representação do modelo unificado proposto. A revisão feita por WEHMUTH et al. (2014a) é descrita a seguir.

2.3.1 Modelos baseados em *snapshots*

Alguns modelos baseados em *snapshots* para GVTs adotam um grafo agregado e uma sequência de sucessivos subgrafos que representam o estado da rede em instantes de tempo discretos. Nesta classe se encontram os modelos propostos por BUI-XUAN et al. (2003) e FERREIRA (2004). Os modelos propostos por HOLME (2005) e HOLME and SARAMAKI (2012) também são baseados em *snapshots*, onde as arestas são representadas como tuplas da forma (i, j, t) , que significam a existência de um contato entre os nós i e j no instante de tempo t . Nesta classe ainda se encontram os modelos propostos por TANG et al. (2009), TANG et al. (2010a), TANG et al. (2010b), bem como todos aqueles quem modelam um GVT como uma sequência de grafos estáticos (*snapshots*) onde cada um deles representa o GVT em um dado instante de tempo.

Modelos baseados em *snapshots* são amplamente usados na literatura e geralmente dão uma noção simples e intuitiva de GVTs. Entretanto, modelos baseados em *snapshots* demandam algumas suposições a serem feitas e que não estão explicitamente implantadas no modelo. Por exemplo, é geralmente assumido em tais modelos que os nós tenham um tipo de memória que permita que a transitividade induzida pelas arestas seja propagada em cada nó ao longo do tempo. Isto significa que um caminho pode ser construído através de um dado nó, mesmo que este nó esteja desconectado de todos os outros durante um período de tempo, o que significa que o nó é capaz de reter a transitividade da aresta durante o período de desconexão. Mesmo que este comportamento seja simples e intuitivamente esperado, em muitos casos, ele não é idealizado junto com o modelo e tem de ser assumido como uma propriedade adicional (externa), o que resulta no fato de que este pressuposto também deva ser incorporado aos algoritmos utilizados com o modelo em questão, tornando tais algoritmos dependentes desses pressupostos externos.

É possível perceber que os modelos baseados em *snapshots* usam somente arestas que conectam os nós presentes em um mesmo instante de tempo. Isto é consistente com o conceito de arestas dinâmicas espaciais proposto no modelo unificado. Desta forma, todos os GVTs modelados por *snapshots* podem ser representados pelo modelo unificado usando apenas arestas dinâmicas espaciais, dado que todos os nós necessários e os instantes de tempo estão presentes nele.

Assim, um GVT com quatro nós e três instantes de tempo na classe de modelos baseados em *snapshots* seria representado no modelo unificado por um GVT cujo tensor

de adjacência em forma de matriz é igual ao apresentado na Figura 1. As entradas que contêm “*” podem ter valores diferentes de zero, indicando a presença potencial de arestas espaciais. Note que todas as entradas deste tipo estão localizadas na diagonal principal de blocos da matriz do tensor adjacência, que na verdade corresponde aos *snapshots*.

$$\text{Mat}(A_{\mathbf{w}}) = \begin{array}{c} \left[\begin{array}{cccc|cccc|cccc} * & * & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ \hline 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \end{array} \right] \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array} \begin{array}{l} t_0 \\ t_1 \\ t_2 \end{array}$$

Figura 1 – Modelo baseado em *snapshots* representado pelo modelo unificado

Segundo WEHMUTH et al. (2014a), qualquer GVT modelado pela classe de *snapshots* pode ser representado de maneira simples no modelo unificado. Na verdade, os modelos baseados em *snapshots* são, de fato, um sub-espaco dos GVTs representáveis no modelo unificado. Além disso, nesses modelos, os *snapshots* estão formalmente desconectados entre si (embora estejam implicitamente ligados), uma vez que as ligações temporais não são explicitamente construídas nesta classe.

2.3.2 Modelos baseados em intervalos de tempo contínuos

Esta classe inclui modelos que usam uma função de presença definida sobre intervalos de tempo contínuos ($t \in R^+$), tais como a versão de tempo contínuo de GVT proposto por CASTEIGTS et al. (2011). Uma função de presença indica se uma determinada aresta está disponível em um determinado instante de tempo. Assume-se ainda que a função de presença usada em tais modelos é construída de tal maneira que cada intervalo de tempo (ou a sua união) tem uma medida finita e diferente de zero. Embora essa suposição não esteja explicitamente indicada no trabalho original, ela é consistente com todos os exemplos e os raciocínios presentes.

É importante dizer que o modelo unificado proposto por WEHMUTH et al. (2014a) é uma versão discreta do modelo do intervalo de tempo contínua, que retém, no entanto, toda a informação presente no modelo original com base em intervalos de tempo contínuo.

O modelo unificado pode representar a classe dos modelos baseados em intervalos de tempo contínuos de diferentes formas. Uma delas é por meio da utilização de arestas dinâmicas mistas. Os instantes de tempo presentes em cada aresta mista (u, t_a, v, t_b) são capazes de representar os intervalos de tempo nos quais a aresta existe $([t_a, t_b])$.

Outra forma é mapear o intervalo de tempo contínuo em *snapshots*. Se uma aresta dinâmica existe entre dois instantes de tempo t_a e t_b , todo *snapshot* compreendido no intervalo $[t_a, t_b]$ pode representar a aresta como uma aresta espacial. Visto que modelos baseados em *snapshots* podem ser representados pelo modelo unificado, os intervalos de tempo contínuo mapeados em *snapshots* também podem.

Uma terceira abordagem é a derivação da forma anterior, onde, além das arestas espaciais, seriam colocadas arestas dinâmicas temporais entre nós temporais nos instantes de tempo sucessivos. Dessa forma, as arestas temporais e espaciais conseguem garantir a conectividade contínua dos nós temporais entre os intervalos de tempo subsequentes.

A representação de um GVT por um modelo baseado em intervalos de tempo contínuos (a) e as diferentes formas de representação do mesmo GVT pelo modelo unificado (b, c, d) se encontram na Figura 2.

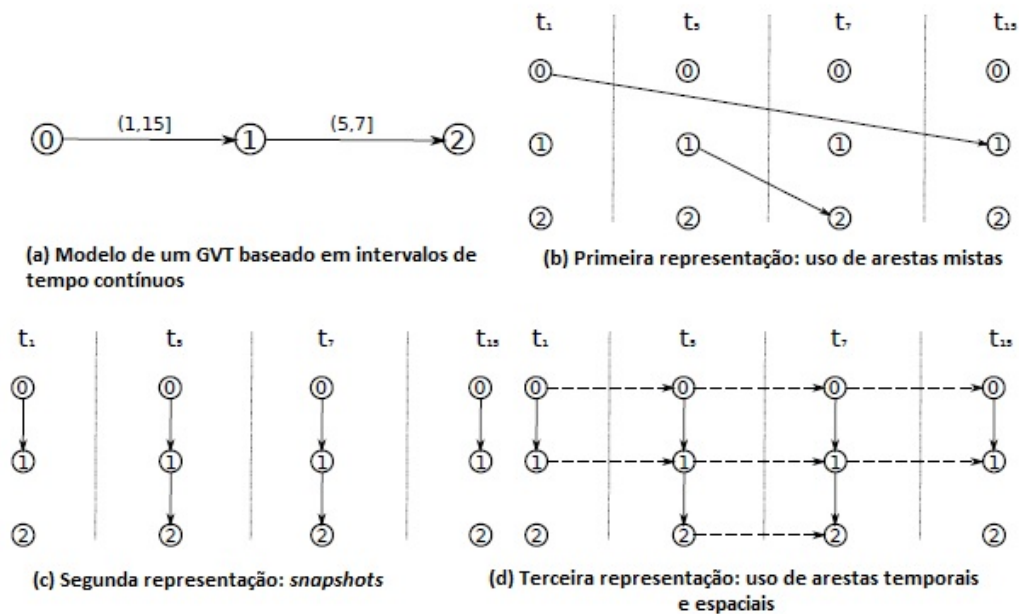


Figura 2 – Conversão de intervalos de tempo contínuos em GVTs discretos

2.3.3 Modelos baseados em arestas espaciais e temporais

Alguns modelos como o proposto por KOSTAKOS (2009) são fundamentados na ideia de que uma classe de conexões representa interações instantâneas entre diferentes nós enquanto outra classe representa um estado de espera de um dado nó. Esses conceitos são formalizados no modelo unificado proposto pelo uso de arestas dinâmicas espaciais e temporais. No modelo unificado, tais arestas dinâmicas são completamente formalizadas e podem ser usadas para fazer uma representação não ambígua desse tipo de GVT. Na figura 3, as entradas diferentes de zero no formato matricial do tensor de adjacência de um GVT demonstram as arestas espaciais (na diagonal principal de blocos) e as arestas temporais (os demais “*”).

$$\text{Mat}(A_{\mathbf{w}}) = \begin{array}{c} \left[\begin{array}{cccc|cccc|cccc} * & * & * & * & * & 0 & 0 & 0 & * & 0 & 0 & 0 \\ * & * & * & * & 0 & * & 0 & 0 & 0 & * & 0 & 0 \\ * & * & * & * & 0 & 0 & * & 0 & 0 & 0 & * & 0 \\ * & * & * & * & 0 & 0 & 0 & * & 0 & 0 & 0 & * \\ \hline 0 & 0 & 0 & 0 & * & * & * & * & * & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & * & * & * & * & 0 & 0 & 0 & * \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ \hline 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \end{array} \right] \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array} \begin{array}{l} t_0 \\ \\ \\ \\ \hline t_1 \\ \\ \\ \\ \hline t_2 \end{array}$$

Figura 3 – Representação de um GVT baseado em arestas espaciais e temporais

2.3.4 Modelos baseados em arestas temporais e mistas

Alguns trabalhos encontrados na literatura recente, como o de KIM and ANDERSON (2012), sugerem o uso de arestas que conectam os nós em diferentes instantes de tempo. GVTs desta classe podem ser representados no modelo unificado utilizando apenas as arestas dinâmicas temporais e mistas. A Figura 4 mostra o formato matricial do tensor de adjacência de um GVT desta classe. Nota-se que um modelo de GVTs baseado em arestas temporais e mistas é um caso particular do modelo unificado proposto por WEHMUTH et al. (2014a).

$$\text{Mat}(A_{\mathbf{w}}) = \begin{array}{c} \left[\begin{array}{cccc|cccc|cccc} 0 & 0 & 0 & 0 & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & * & * & * \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 & 0 & 1 & 2 & 3 \end{array} \right] \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \\ t_0 \\ 0 \\ 1 \\ 2 \\ 3 \\ t_1 \\ 0 \\ 1 \\ 2 \\ 3 \\ t_2 \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array}$$

Figura 4 – Representação de um GVT baseado em arestas temporais e mistas

3 MODELO DE REPRESENTAÇÃO UNIFICADO DE GVTs

O modelo de representação de GVTs utilizado neste trabalho, bem como sua formalização, descrita a seguir, são encontrados em WEHMUTH et al. (2014a). Segundo os próprios autores, o modelo criado é suficientemente geral para capturar as necessidades de redes dinâmicas distintas, de tal forma que ele consegue representar várias classes de modelos para redes dinâmicas encontradas na literatura recente. Tais classes de modelos geralmente não são capazes de representar umas as outras.

O modelo representa um GVT como um objeto $H = (V, E, T)$, onde V é o conjunto de nós, T é o conjunto finito de instantes de tempo para os quais o GVT é definido, e $E \subseteq V \times T \times V \times T$ é o conjunto de arestas. Consoante à notação adotada em WEHMUTH et al. (2014a), $V(H)$ representa o conjunto de todos os nós em H , $E(H)$ o conjunto de todas as arestas em H e $T(H)$ o conjunto de todos os instantes de tempo em H . Além disso, $n(H) = |V(H)|$ corresponde ao número de nós em H , $m(H) = |E(H)|$ o número de arestas em H e $\tau(H) = |T(H)|$ o número de instantes de tempo no qual H é definido.

Uma aresta dinâmica e em um GVT H é definida como uma quadrupla ordenada $e = (u, t_a, v, t_b)$, onde $u, v \in V(H)$ são, respectivamente, nós de origem e destino, enquanto que $t_a, t_b \in T(H)$ são os instantes de tempo de origem e de destino. É possível que u e v sejam iguais, da mesma forma que pode ocorrer com t_a e t_b . A aresta dinâmica $e = (u, t_a, v, t_b)$ pode ser entendida como uma conexão do nó u no tempo t_a para o nó v no tempo t_b . De acordo com a natureza temporal das ligações, uma aresta dinâmica é uma aresta direcionada; entretanto, se houver a necessidade de representar arestas não orientadas em um GVT, a aresta (u, t_a, v, t_b) e sua recíproca (v, t_b, u, t_a) devem estar presentes em $E(H)$.

Ainda em WEHMUTH et al. (2014a), foram definidas quatro projeções canônicas, onde cada projeção mapeia uma aresta dinâmica em cada um dos seus componentes:

$$\begin{aligned} \pi_1 : E(H) &\longrightarrow V(H) \\ &(u, t_a, v, t_b) \mapsto u, \\ \pi_2 : E(H) &\longrightarrow T(H) \\ &(u, t_a, v, t_b) \mapsto t_a, \\ \pi_3 : E(H) &\longrightarrow V(H) \\ &(u, t_a, v, t_b) \mapsto v, \\ \pi_4 : E(H) &\longrightarrow T(H) \\ &(u, t_a, v, t_b) \mapsto t_b. \end{aligned}$$

Uma aresta dinâmica genérica $e = (u, t_a, v, t_b)$ pode ser classificada de acordo com

quatro classes, dependendo de suas características temporais:

1. Arestas Espaciais conectam dois nós em um mesmo instante de tempo, e está na forma $e = (u, t_a, v, t_a)$, onde $u \neq v$;
2. Arestas Temporais conectam um mesmo nó em dois instantes de tempo distintos, e está na forma $e = (u, t_a, u, t_b)$, onde $t_a \neq t_b$;
3. Arestas Mistas conectam diferentes nós em diferentes instantes de tempo, e está na forma $e = (u, t_a, v, t_b)$, onde $u \neq v$ e $t_a \neq t_b$;
4. Arestas *Self-Loop* Espaço-Temporais conectam um mesmo nó no mesmo instante de tempo, e está na forma $e = (u, t_a, u, t_a)$.

Um nó temporal é representado pelo par ordenado (u, t_a) , onde $u \in V(H)$ é um nó e $t_a \in T(H)$ é um instante de tempo. Um nó temporal é a definição de um dado nó em um dado instante de tempo. O conjunto de nós temporais em um GVT H é dado por $V(H) \times T(H)$, o produto cartesiano do conjunto de nós e o conjunto de instantes de tempo. O conjunto de todos os nós temporais de um GVT H é denotado por $VT(H) = V(H) \times T(H)$. As projeções canônicas $\pi_1(u, t_a, v, t_b)$ e $\pi_2(u, t_a, v, t_b)$ extraem o nó e o instante de tempo que compõem o nó temporal. A notação (u, \bullet) representa a presença do nó u em qualquer instante de tempo.

A definição de um GVT H é a mais geral possível e não impõe qualquer ordem no conjunto de tempo $T(H)$. Contudo, é utilizada, normalmente, a noção clássica de tempo, que impõe uma ordem total sobre $T(H)$. Dentro de tal contexto, onde $T(H)$ tem uma ordem linear, tanto arestas mistas ou temporais $e = (u, t_a, v, t_b)$ também podem ser classificadas como progressivas ou regressivas, dependendo da ordem de seus componentes temporais. Arestas dinâmicas que são originadas em um instante de tempo anterior e destinadas a um instante de tempo mais tarde, são progressivas ($t_a < t_b$), enquanto que as arestas dinâmicas originadas em um instante de tempo posterior e destinadas a um instante de tempo anterior são regressivas ($t_a > t_b$). Arestas regressivas são particularmente úteis para a criação de GVTs cíclicos, que por sua vez podem ser aplicados para modelar redes com um comportamento periódico cíclico.

É importante salientar que o modelo de representação de GVTs em questão é um caso particular de grafo multi-aspectos (WEHMUTH et al., 2014b), onde o aspecto principal em discussão é o tempo. Um grafo multi-aspecto é capaz de representar múltiplas camadas e grafos variantes no tempo. Neste âmbito, conjunto de vértices, instantes de tempo ou quaisquer outras características independentes são considerados aspectos.

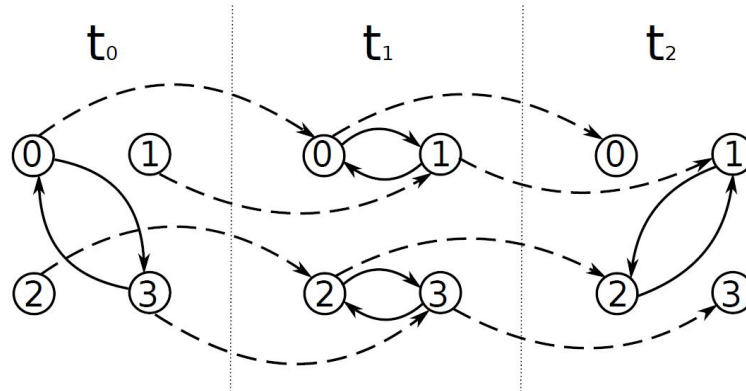


Figura 5 – Exemplo para a representação de um GVT

3.1 REPRESENTAÇÃO ESTRUTURAL DE UM GVT

Considere o GVT W apresentado na figura 5. Ele será utilizado ao longo desta seção para exemplificar o modelo de representação unificado de GVTs adotado neste trabalho e a classificação das arestas dinâmicas. Formalmente, $W = (V, E, T)$, onde V é o conjunto de nós, T é o conjunto finito de instantes de tempo e E é o conjunto de arestas dinâmicas não ponderadas (peso igual a 1). Por definição, $V(W)$ representa o conjunto de todos os nós em W , $E(W)$ o conjunto de todas as arestas em W e $T(W)$ o conjunto de todos os instantes de tempo em W , tem-se que:

- $V(W) = \{0, 1, 2, 3\}$. Logo, $|V(W)| = 4$
- $T(W) = \{t_0, t_1, t_2\}$. Logo, $|T(W)| = 3$
- $E(W) = \{\forall e = (u, t_a, v, t_b), \text{ onde } u, v \in V(W) \text{ e } t_a, t_b \in T(W)\}$

Dentre as arestas dinâmicas mostradas na figura 5, as arestas espaciais são representadas pelas setas sólidas e as arestas temporais pelas setas tracejadas. Não há arestas mistas e nem arestas *self-loop* espaço-temporais retratadas no GVT W .

Na subseção 3.1.1, é abordada a representação algébrica de um GVT por meio de tensores de adjacência. Na subseção 3.1.2, é discutida a representação matricial do tensor de adjacência de um GVT, usando o GVT W como exemplo.

3.1.1 Tensor de Adjacência de um GVT

Um tensor é uma entidade geométrica usada para descrever propriedades físicas, que acaba por generalizar, de fato, a noção de escalares e vetores. Um escalar é definido por um tensor de ordem zero, e representado por um número; um vetor, é definido por um tensor de ordem um, e representado por uma matriz linha/coluna. De uma forma mais ampla, um tensor de ordem n pode ser representado por uma matriz de dimensão n .

De acordo com o modelo de representação unificado de GVTs utilizado nesse trabalho, a estrutura das arestas dinâmicas é uma quádrupla ordenada da forma $e = (u, t_a, v, t_b)$. Dessa forma, pode-se definir que o tensor de adjacência de um GVT H é um tensor de quarta ordem $A(H)$ com dimensão $|V(H)| \times |T(H)| \times |V(H)| \times |T(H)|$ que tem uma entrada para toda possível aresta dinâmica em H (WEHMUTH et al., 2014a).

Cada aresta dinâmica presente no GVT H corresponde a um valor diferente de zero no tensor de adjacência $A(H)$. Tais valores podem representar o peso de uma aresta dinâmica no caso de GVTs ponderados ou podem ser iguais a 1 em GVTs não ponderados.

3.1.2 Representação Matricial do Tensor de Adjacência de um GVT

Conforme demonstrado em WEHMUTH et al. (2014a), para cada GVT H com n nós e τ instantes de tempo, onde $n > 1$ e $\tau > 1$, existe um grafo estático orientado G com $n \times \tau$ nós que é isomorfo a H . Dois grafos $G = (V_1, E_1)$ e $H = (V_2, E_2)$ são ditos isomorfos se existe uma função bijetora entre vértices de G para vértices de H , tal que suas adjacências sejam preservadas. Dessa forma, a matriz do tensor de adjacência de um GVT pode ser representada pela matriz de adjacência do grafo orientado isomorfo correspondente.

Contudo, é necessário compreender que o grafo estático orientado é obtido por meio da representação direta dos nós temporais de um GVT: cada nó temporal (u, t_a) no GVT equivale a um nó no grafo estático. Assim, o grafo estático tem $|V| \times |T|$ nós, e, conseqüentemente, sua matriz de adjacência tem $|V| \times |T| \times |V| \times |T| = |V|^2 \times |T|^2$ entradas. Cada entrada diferente de zero nessa matriz corresponde a uma aresta dinâmica do GVT.

Considere o GVT W que serve de exemplo para a compreensão do modelo de representação unificado apresentado. A figura 6 mostra a forma matricial do tensor de adjacência do GVT W , $Mat(A_W)$. Cada um dos quatro nós do GVT W é identificado por um número (0, 1, 2 e 3). Os três instantes de tempo são identificados sequencialmente por t_0, t_1, t_2 .

É possível observar que $Mat(A_W)$ é dividida em blocos temporais. Cada bloco temporal representa o conjunto ordenado de arestas dinâmicas existentes (ou ausentes) entre nós genéricos u e v , onde $u, v \in V$, do tempo t_a e para o tempo t_b , onde $t_a, t_b \in T$, conforme a definição para arestas dinâmicas citada anteriormente.

Deste modo, ao considerar que a matriz $Mat(A_W)$ pode ser expressa também em blocos temporais, pode-se concluir que o número de entradas de $Mat(A_W)$ é igual a $|T| \times |T|$, onde cada entrada equivale a um bloco temporal que retrata todas as arestas dinâmicas entre o tempo de origem t_a e o tempo de destino t_b .

A interpretação dos valores da matriz $Mat(A_W)$ depende da classificação das características temporais de cada aresta dinâmica. A representação de cada uma das

$$\text{Mat}(\mathbf{A}_W) = \begin{array}{c} \left[\begin{array}{cccc|cccc|cccc} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right] \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \\ \hline 0 \\ 1 \\ 2 \\ 3 \end{array} \end{array} \begin{array}{l} \mathbf{t}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \end{array}$$

Figura 6 – Forma matricial do tensor de adjacência do GVT W (WEHMUTH et al., 2014a)

quatro classes de arestas dinâmicas é apresentada a seguir:

- *Arestas Espaciais.* Os blocos temporais que estão na diagonal principal da matriz $\text{Mat}(A_W)$ (levemente sombreados na figura 6) equivalem às arestas espaciais existentes em um determinado instante de tempo. Isso significa que, todo bloco temporal representado em $\text{Mat}(A_W)$, cujos tempos de origem e destino são os mesmos ($t_a = t_b$), e, por consequência, as arestas dinâmicas de forma (u, t_a, v, t_a) , representam o estado da rede naquele instante de tempo t_a . Então, pode-se concluir que os blocos temporais na diagonal principal de $\text{Mat}(A_W)$ correspondem à uma sequência ordenada de *snapshots* que retrata o estado da rede em cada instante de tempo e sua respectiva evolução.
- *Arestas Temporais.* Uma aresta temporal é aquela que conecta um mesmo nó u em dois instantes de tempo distintos t_a e t_b . Sendo assim, as arestas temporais estão representadas na diagonal principal dos blocos que estão fora da diagonal principal da matriz $\text{Mat}(A_W)$. Isso está em conformidade a definição formal de arestas temporais, uma quadrupla ordenada (u, t_a, u, t_b) , com $t_a \neq t_b$. As arestas temporais estão retratadas na figura 6, nos campos com fundo branco na matriz $\text{Mat}(A_W)$.
- *Arestas Mistas.* As arestas mistas são aquelas que conectam diferentes nós u e v em diferentes instantes de tempo t_a e t_b . As arestas mistas estão representadas fora da diagonal principal dos blocos que estão fora da diagonal principal da matriz $\text{Mat}(A_W)$. Na figura 6, as posições das arestas mistas são retratadas nos campos

com o fundo mais escuro; entretanto, não há arestas mistas representadas no GVT W .

- *Arestas Self-Loop Espaço-Temporais.* Dependendo do modelo de representação de GVTs, é possível reproduzir uma aresta que conecta um nó arbitrário u a ele mesmo, e em um mesmo instante de tempo t_a . Esses tipos de arestas podem ser representadas nos campos presentes na diagonal principal de qualquer bloco que esteja na diagonal principal da matriz $Mat(A_W)$. No GVT W e na figura 6, não é retratado esse tipo de aresta.

4 TRABALHOS RELACIONADOS

Um considerável número de trabalhos em redes dinâmicas preocupam-se essencialmente com as estruturas de dados necessárias para modelar GVTs e com as definições de centralidade de nó. Como GVTs diferem da representação usual de grafos, modelos tradicionais e definições podem não refletir fielmente a evolução do estado da rede.

Este capítulo apresenta um conjunto de trabalhos encontrados na literatura referentes a métricas de centralidade em redes temporais. Além disso, na seção 4.1, são mostrados os modelos mais comuns de difusão de informação em redes: modelos epidêmicos e *Random Walk*.

Até onde sabemos, este é o primeiro trabalho que aborda a centralidade de tempo em GVTs. De fato, os trabalhos anteriores estão mais interessados em qualificar a importância dos nós em uma rede do que em determinar os instantes de tempo mais importantes no contexto de um GVT. Diferente desses trabalhos, a centralidade de tempo visa qualificar a importância dos instantes de tempo, a qual pode ser particularmente útil para definir melhores momentos para difusão, gerar modelos e prever o comportamento de GVTs.

Essencialmente, as métricas de centralidade tradicionais foram desenvolvidas para avaliar a importância dos nós em grafos estáticos. Portanto, elas podem não ser aplicadas diretamente em GVTs. Entretanto, existem trabalhos na literatura que tentam adaptar as medidas de centralidade para redes dinâmicas. Em muitos deles, um GVT é considerado como uma sequência ordenada de grafos estáticos que representam a variação do estado da rede ao longo do tempo.

KIM and ANDERSON (2012) apresentam um modelo chamado *time-ordered graph*, que transforma redes dinâmicas em redes estáticas. A ideia central é modelar uma rede dinâmica como um digrafo ordenado pelo tempo que associa os *snapshots* temporais com as arestas orientadas que ligam cada nó com seu sucessor no tempo. Sendo assim, algoritmos desenvolvidos para redes estáticas e variações das métricas tradicionais de centralidade podem ser aplicadas, sob aspectos temporais, para redes dinâmicas.

De maneira independente, NICOSIA et al. (2013); TANG et al. (2010a) também propõem uma redefinição ou extensão apropriada das métricas tradicionais para GVTs. Neste caso, o conceito a ser estendido é o de caminho mínimo, que agora deve considerar não somente uma distância geodésica entre os nós, mas também uma distância temporal.

A seguir, são discutidas as principais métricas de centralidade de nó, no ambiente de GVTs. Diferentes definições de *Betweenness* Temporal e *Closeness* Temporal são encontradas nos trabalhos de NICOSIA et al. (2013), KIM and ANDERSON (2012), TANG et al. (2010a) e uma definição de Grau Temporal é mostrada por KIM and ANDERSON (2012):

- Grau temporal

Formalmente, o modelo *time-ordered graph* consiste em uma rede dinâmica $G_{[0,T]}^D = (V, E_{0,T})$, onde $[0, T]$ é o intervalo de tempo entre os instantes de tempo inicial e final, V é o conjunto de nós e $E_{0,T}$ é o conjunto de arestas temporais nas quais uma aresta temporal $(u, v)_{i,j} \in E_{0,T}$ existe entre os nós u e v no intervalo de tempo $[i, j]$, com $i \leq T$ e $j \geq 0$.

O grau temporal $D_{i,j}(v)$ de um nó $v \in V$ em um intervalo de tempo $[i, j]$ é dado pelo número total normalizado de arestas de entrada em v e arestas de saída de v no intervalo de tempo $[i, j]$, desconsiderando arestas *self-loop* entre v_{t-1} e v_t para todo $t \in \{i+1, \dots, j\}$. Sendo assim, o grau temporal $D_{i,j}(v)$ é igual a $\sum_{t=i}^j 2D_t(v)$, onde $D_t(v)$ é o grau de v em G_t , e sua normalização pode ser obtida pela divisão por $2(|V| - 1)m$, onde $m = j - i$.

- *Betweenness* temporal

Conforme NICOSIA et al. (2013) e TANG et al. (2010a), uma maneira simples de estender a centralidade de *betweenness* para GVTs consiste na contagem da fração de caminhos mínimos temporais que atravessam um determinado nó. Entretanto, pode ser importante levar em consideração não somente o número de caminhos mínimos temporais que passam por um nó, mas também por quanto tempo um nó, ao longo do caminho mais curto, retém uma mensagem antes de repassá-la para o próximo nó.

Considere um caso onde os nós i e j são conectados por apenas um único caminho mínimo que consiste de duas arestas $(i, k)_{tl}$ e $(k, j)_{tm}$. Isso significa que a aresta que conecta i a k aparece no tempo tl , enquanto a aresta que conecta k a j aparece no tempo tm . Uma vez que o caminho que atravessa k é o único para i alcançar j temporalmente, então pode-se dizer que k tem um papel central na comunicação entre i e j . No entanto, a vulnerabilidade do nó k depende do intervalo $[tl, tm]$: quanto maior este intervalo de tempo, maior a probabilidade de que uma mensagem encaminhada para k se perca caso k seja removido da rede. A fim de considerar o efeito dos tempos de espera, a centralidade de *betweenness* temporal do nó i no tempo tm é definida como:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j \neq i} \sum_{k \neq j, k \neq i} \frac{U(i, tm, j, k)}{\sigma_{jk}},$$

onde σ_{jk} é o número de caminhos mínimos temporais de j para k , e $U(i, tm, j, k)$ é o número de caminhos temporais mínimos de i para j nos quais o nó i é atravessado a partir do caminho no *snapshot* tm ou em um *snapshot* anterior $t' < tm$, de modo que a próxima aresta do mesmo caminho seja avaliada em um *snapshot* posterior $t'' > tm$.

Segundo KIM and ANDERSON (2012), o *betweenness* temporal $B_{i,j}(v)$ de um nó $v \in V$ em um intervalo de tempo $[i, j]$ deve ser a proporção de todos os menores caminhos temporais que passam por v pelo número total de caminhos mínimos temporais sobre todos os pares de nós para cada intervalo de tempo em $\{[t, j] : i \leq t < j\}$.

Seja $S_{x,y}(s, d)$ o conjunto de caminhos mínimos temporais a partir de um nó de origem s e um nó de destino d em um intervalo de tempo $[x, y]$ e $S_{x,y}(s, d, v)$ o subconjunto de $S_{x,y}(s, d)$ que consiste de caminhos que contêm o nó v em seu interior. Dessa forma, o *betweenness* temporal de um nó v é:

$$B_{i,j}(v) = \sum_{i \leq t < j} \sum_{s \neq v \neq d \in V} \frac{\sigma_{t,j}(s, d, v)}{\sigma_{t,j}(s, d)},$$

onde $\sigma_{t,j}(s, d) \equiv |S_{t,j}(s, d)|$ e $\sigma_{t,j}(s, d, v) \equiv |S_{t,j}(s, d, v)|$.

- *Closeness* temporal

De acordo com TANG et al. (2010a), dois nós de um grafo estático são ditos próximos um do outro se sua distância geodésica é pequena. Em um grafo estático, uma estimativa do *closeness* de um nó i é obtida pelo comprimento médio do menor caminho mínimo estático para todos os outros nós no grafo. De maneira similar, pode-se estender a definição de *closeness* para redes temporais usando o comprimento do caminho mínimo temporal entre os nós, que é a medida de quanto mais cedo um nó de origem pode enviar uma mensagem para todos os outros nós (PAN and SARAMAKI, 2011). Dada a menor distância temporal $d_{ij}(t_{min}, t_{max})$, a centralidade de *closeness* temporal pode ser expressa por:

$$C_i^h = \frac{1}{W(N-1)} \sum_{j \neq i \in V} d_{i,j}^h,$$

de modo que os nós que têm, em média, menores distâncias temporais para outros nós, são considerados mais centrais.

Já segundo KIM and ANDERSON (2012), o *closeness* temporal $C_{i,j}(v)$ de um nó $v \in V$ em um intervalo de tempo $[i, j]$ é a inversa da soma das distâncias do caminhos mínimos temporais para todos os outros nós em $V \setminus v$ em cada intervalo de tempo em $\{[t, j] : i \leq t < j\}$. São considerados $m = j - i$ intervalos de tempo, onde o tempo inicial t de cada intervalo de tempo varia de i até $j - 1$. Formalmente, o *closeness* temporal de um nó v é:

$$C_{i,j}(v) = \sum_{i \leq t < j} \sum_{u \in V \setminus v} \frac{1}{\Delta_{t,j}(v, u)},$$

onde $\Delta_{t,j}(v, u)$ é a distância do caminho mínimo temporal a partir de v até u no intervalo de tempo $[t, j]$. Se não existe caminho temporal de v até u no intervalo de tempo $[t, j]$, $\Delta_{t,j}(v, u)$ é definido como ∞ .

Finalmente, é importante dizer que existe uma diferença primordial entre a centralidade de tempo, proposta neste trabalho, e as métricas temporais de centralidade existentes na literatura. Enquanto as métricas temporais visam avaliar a importância relativa dos nós em uma rede dinâmica ou GVT (assim como as métricas de centralidade tradicionais em grafos estáticos), a centralidade de tempo pretende avaliar a importância de um instante de tempo frente aos demais instantes de tempo em um GVT.

4.1 MODELOS DE DIFUSÃO DE INFORMAÇÃO EM REDES

O processo de difusão de informação é comum a diferentes áreas do conhecimento, além da computação e das redes de computadores (KHELIL et al., 2002), como na epidemiologia (STEHL et al., 2011a; SHULGIN et al., 1998) e em ciências sociais (STEHL et al., 2011b; BARTHOLOMEW, 1973).

Em suma, modelos de difusão de informação podem ser divididos em duas classes: os modelos epidêmicos e o modelo *Random Walk*. EASLEY and KLEINBERG (2010) fazem uma revisão sobre os modelos epidêmicos mais básicos e BARRAT et al. (2008); STARNINI et al. (2012) mostram a dinâmica da transmissão de informação em redes usando o modelo *Random Walk*.

4.1.1 Modelos Epidêmicos

Conforme EASLEY and KLEINBERG (2010), os padrões pelos quais epidemias se espalham através de grupos de pessoas são determinados não apenas pelas propriedades do patógeno - contágiosidade, duração do período infeccioso e sua gravidade - mas também através das estruturas da rede dentro da população afetada. A rede social dentro de uma população - indivíduos conhecem outros indivíduos - determina muito sobre como é susceptível a transmissão de uma doença de uma pessoa para outra. Em termos mais gerais, as oportunidades para que uma doença se espalhe são dadas por uma rede de contatos: existe um nó para cada pessoa, e uma aresta, se duas pessoas entram em contato entre si, de uma forma que torna possível para a transmissão da doença de uma para a outra.

Dessa forma, é possível observar que a propagação de doenças, rumores ou ideias em redes sociais e difusão de informação em redes dinâmicas têm, essencialmente, as mesmas características. Conseqüentemente, os modelos existentes na literatura que descrevem

o comportamento de transmissão de doenças também são adequados para a difusão de informação em redes dinâmicas.

Ainda segundo EASLEY and KLEINBERG (2010), o modelo mais simples de contágio é conhecido como processo de ramificação (*Branching process*). Esta é sua descrição:

- *Primeira onda.* Suponha que uma pessoa que traz consigo uma nova doença entra em uma população, e transmite-a a cada pessoa que ela conhece de forma independente com uma probabilidade p . Além disso, suponha que ela conheça k pessoas durante o período no qual ela está infectada; essas k pessoas são a primeira onda da epidemia. Com base na transmissão probabilística da doença, algumas das pessoas na primeira onda podem ser infectadas com a doença, enquanto outras não.
- *Segunda onda.* Cada pessoa na primeira onda vai para a população e encontra k pessoas diferentes, resultando em uma segunda onda de $k \times k = k^2$ pessoas. Cada pessoa infectada na primeira onda passa a doença de forma independente para cada uma das k pessoas na segunda onda, mais uma vez de forma independente, com probabilidade p .
- *Ondas subsequentes.* Novas ondas são formadas da mesma maneira, fazendo com que cada pessoa na onda atual encontre k novas pessoas, transmitindo a doença independentemente, com uma probabilidade p .

Formalmente, dado um GVT $G = (V, E, T)$ que representa uma rede de contatos ao longo do tempo, onde V é o conjunto de nós, E o conjunto de contatos entre os nós e T é o conjunto de instantes de tempo, se existe contato entre dois nós $u, v \in V$, e u se tornou infectado em algum instante de tempo, a doença/informação pode atingir v , com probabilidade p , em um instante de tempo futuro.

O processo de ramificação é um modelo simplificado de difusão de informação/doença. Existem modelos na literatura que são capazes de representar redes de contato mais complexas, como os modelos epidêmicos SIR e SIS.

4.1.1.1 Modelo Epidêmico SIR

Um nó individual no modelo de processo de ramificação passa por três fases potenciais durante o curso da epidemia:

- *Suscetível (S):* antes de um nó ser atingido pela doença/informação, ele está suscetível à infecção a partir de seus vizinhos.
- *Infectado (I):* uma vez que o nó foi atingido pela doença/informação, ele é considerado infectado e tem uma probabilidade de infectar cada um dos seus vizinhos.

- *Removido (R)*: depois que um nó tenha passa por um período de infecção completo, o nó é considerado removido do processo de ramificação, e já não representa uma ameaça de infecção futura.

Cada nó tem o potencial de estar em um desses três estados. O progresso da epidemia é controlado pela estrutura da rede de contatos e por duas grandezas adicionais: p (a probabilidade de infecção) e t_I (o comprimento da infecção). Inicialmente, um ou mais nós estão no estado I e todos os outros estão no estado S . Cada nó u que entra no estado I permanece infectado por um certo período de tempo ou número de passos t_I . Durante o período de infecção, v tem uma probabilidade p de transmitir a doença/informação para todos os seus vizinhos que estão no estado S . Depois de t_I passos, o nó u não está mais infectado ou suscetível para novas infecções; então ele é considerado removido, já que é agora um nó inerte na rede de contatos que não pode quer adquirir ou transmitir a doença/informação.

4.1.1.2 Modelo Epidêmico SIS

Neste modelo, os nós podem alternar apenas entre dois possíveis estados: Suscetível (S) e Infectado (I). Não há o estado Removido; pelo contrário, depois que um nó conclui seu período de infecção, ele volta ao estado S e está pronto para adquirir a doença/informação novamente.

O mecanismo do modelo SIS se dá da seguinte forma: inicialmente, um ou mais nós estão no estado I e todos os demais no estado S . Cada nó u que entra no estado I permanece infectado por um certo período de tempo ou número de passos t_I . Durante o período de infecção, v tem uma probabilidade p de transmitir a doença/informação para cada um dos seus vizinhos que estão no estado S . Após t_I passos, o nó v não está mais infectado e, assim, retorna ao estado S .

4.1.2 Modelo *Random Walk*

De acordo com BARRAT et al. (2008); STARNINI et al. (2012); LOVASZ (1996), *Random Walk* (passeio aleatório) é certamente o modelo de difusão mais simples, e sua dinâmica provê sugestões fundamentais para entender toda a classe de processos de difusão em redes. O processo *Random Walk* (RW) é definido por um caminhante (*walker*) que, localizado em um dado nó i no tempo t , salta para o vizinho mais próximo j no tempo $t + 1$.

Considere um GVT $G = (V, E, T)$ que representa uma rede de contatos ao longo do tempo, onde V é o conjunto de nós, E o conjunto de contatos entre os nós e T é o conjunto de instantes de tempo. Ainda, considere a matriz de adjacência A_G , tal que cada

entrada a_{ij} dessa matriz é igual a 1, caso o nó j seja vizinho do nó i , e zero, caso contrário. Assim, a probabilidade de transição a partir de i para j em cada instante de tempo é:

$$p_b = (i \rightarrow j) = \frac{a_{ij}}{\sum_r a_{ir}} \equiv \frac{a_{ij}}{k_i},$$

onde $k_i = \sum_j a_{ij}$ é o grau do vértice i : o caminhante presente em i no tempo t salta para um dos seus vizinhos mais próximos, escolhido uniforme e aleatoriamente entre seus k_i vizinhos, portanto, com probabilidade $1/k_i$. Se o nó i estiver isolado no tempo t (ou seja, $V_i(t) = \emptyset$), o caminhante permanece no nó i . Em todo caso, o tempo é incrementado ($t \rightarrow t + 1$). Neste contexto, a difusão da informação é realizada pela transição do caminhante entre os nós ao longo do tempo.

Ainda segundo STARNINI et al. (2012), outra propriedade característica de *random walk*, relevante em termos de busca em redes, é a noção de cobertura. A cobertura $C(t)$ é definida como o número de diferentes nós que foram visitados pelo caminhante no tempo t , e sua média é calculada para diferentes *random walks* a partir de diferentes fontes. A cobertura pode, assim, ser interpretada como a eficiência de pesquisa da rede, medindo o número de diferentes indivíduos que podem ser alcançados a partir de uma origem arbitrária em um determinado número de passos ou intervalos de tempo.

5 CENTRALIDADE DE TEMPO

Como dito anteriormente, diversos sistemas reais podem ser representados por grafos e redes. Na análise de redes e na teoria de grafos, é comum a tentativa de avaliar a importância dos nós baseado na estrutura da rede e na disposição das ligações entre eles. Tal importância pode ser obtida por meio de métricas de centralidades, como Grau (FREEMAN, 1978), *Betweenness* (FREEMAN, 1977) e *Closeness* (SABIDUSSI, 1966), amplamente mencionadas na literatura. Entretanto, estas medidas de centralidade são fundamentadas apenas pela importância dos nós entre si.

Uma vez que o tempo é uma característica fundamental de sistemas modelados por GVTs, é mais importante identificar os instantes de tempo mais importantes no contexto de um GVT do que identificar os nós mais centrais. Pode ser mais interessante, por exemplo, determinar o momento em que uma vacina deverá ser aplicada em uma população para obter melhores respostas de imunização ou designar o melhor tempo para iniciar uma difusão de informação mais rápida e com menores custos.

Neste trabalho, é definida a *Centralidade de Tempo* em GVTs. Similar às medidas de centralidade de nó, a finalidade da centralidade de tempo é avaliar a importância relativa de um dado instante de tempo comparado com outros instantes de tempo de um mesmo GVT. Assim como acontece com as medidas de centralidade de nó, diferentes definições de centralidade de tempo podem surgir, levando em consideração diferentes noções de importância para instantes de tempo.

A dinâmica da rede é expressa pelo conjunto finito e sequencial de instantes de tempo descrito pelo modelo de representação unificado de GVTs. Tais instantes de tempo podem ser representados por *snapshots*, cuja peculiaridade é retratar o estado pontual das conexões e dos nós em um determinado período de tempo.

É imprescindível dizer que o modelo unificado utilizado neste trabalho é capaz de representar em um *snapshot* não somente as arestas espaciais, mas também a incidência de arestas temporais e mistas. Dessa forma, o estado da rede em um determinado instante de tempo t é influenciado tanto pelas conexões iniciadas e encerradas em t (arestas espaciais e *self-loop*), quanto pelas conexões que se iniciam ou se encerram, exclusivamente, em t (arestas temporais e mistas).

5.1 MÉTRICAS DESENVOLVIDAS PARA CENTRALIDADE DE TEMPO

Nessa seção, são apresentadas as três métricas que foram desenvolvidas para calcular a centralidade de tempo de um determinado instante de tempo em um GVT. São elas: *Grau do Tempo*, *Tempo de Cobertura* e *Centralidade de Cobertura*.

Essencialmente, as métricas estão fundamentadas sob dois fatores: grau dos nós

temporais e difusão de informação em uma rede. Grau do Tempo aprecia a disposição dos nós temporais e a natureza das ligações entre eles em cada instante de tempo, enquanto Centralidade de Cobertura e Tempo de Cobertura consideram o tamanho da rede atingido por uma informação e o tempo de difusão da informação para uma parcela (ou para a totalidade) dos nós de uma rede, respectivamente.

É importante salientar que as métricas de Tempo de Cobertura e Centralidade de Cobertura estão associadas a um problema bem específico, que é a difusão de informação/doença/comportamento como algo probabilístico. Neste trabalho, a fim de explorar um caso extremo de difusão, a probabilidade de infecção p foi fixada em 1, ou seja, o nó infectado transmite a informação para todos os k nós adjacentes no próximo passo do processo de ramificação da infecção. Além disso, foi definido que a duração da infecção t_I é ilimitada, ou seja, depois de infectado pela informação, um nó permanece infectado enquanto o processo de contágio durar, sem que haja possibilidade de que ele seja removido do processo ou seja considerado novamente suscetível à infecção.

O modelo de representação de GVTs adotado neste trabalho foi proposto por WEHMUTH et al. (2014a). Dessa forma, para exemplificar as métricas desenvolvidas para avaliar a centralidade de tempo, ao longo dessa seção será considerado um GVT G genérico, elaborado de acordo com o modelo em questão. Formalmente, $G = (V, E, T)$, onde V é o conjunto finito de nós, T é o conjunto finito de instantes de tempo e E é o conjunto de arestas dinâmicas não ponderadas. Consequentemente, $V(G)$ representa o conjunto de todos os nós em G , $E(G)$ o conjunto de todas as arestas em G , e $T(G)$ o conjunto de todos os instantes de tempo em G . Seja o tamanho do conjunto $T(G)$, $|T(G)|$, igual a N . Cada instante de tempo em $T(G)$ pode ser representado por t_i , com $0 \leq i < N$, de tal forma que o conjunto de instantes de tempo sequenciais pode ser descrito por $T(G) = \{t_0, t_1, \dots, t_{N-1}\}$.

As métricas de Grau do Tempo, Tempo de Cobertura e Centralidade de Cobertura são formalizadas e exemplificadas nas subseções 5.1.1, 5.1.2 e 5.1.3, respectivamente.

5.1.1 Grau do Tempo

Seja um nó temporal genérico (u, t) , onde o nó $u \in V(G)$ e o instante de tempo $t \in T(G)$. Pela definição formal de grau de um nó, e de acordo com as projeções canônicas que mapeiam as arestas dinâmicas, definidas pelo modelo de representação unificado de GVTs, os graus de saída e entrada do nó temporal (u, t) são dados, respectivamente, por:

$$\begin{aligned} \text{deg}^+((u, t)) &= |\{e \in E(G) \mid \pi_1(e) = u, \pi_2(e) = t\}|, \\ \text{deg}^-((u, t)) &= |\{e \in E(G) \mid \pi_3(e) = u, \pi_4(e) = t\}|. \end{aligned}$$

Nesse contexto, $deg^+((u, t))$ é o número de arestas dinâmicas originadas a partir do nó u no instante de tempo t , ao passo que $deg^-((u, t))$ é o número de arestas dinâmicas destinadas ao nó u no instante de tempo t (WEHMUTH et al., 2014a). Dessa forma, pode-se concluir que o grau do nó temporal (u, t) é dado por:

$$deg((u, t)) = deg^+((u, t)) + deg^-((u, t))$$

Observe que esta descrição ressalta que o cálculo do grau de um nó temporal compreende todas as quatro classes de arestas dinâmicas definidas pelo modelo de representação unificado. Sendo assim, um determinado instante de tempo t retrata a disposição das conexões iniciadas e terminadas entre nós temporais da forma (u, t) .

Por definição, a centralidade de grau do tempo t_i , $TDC(t_i)$, é dada pela soma dos graus dos nós temporais (u, t_i) , onde u se refere a todo o nó pertencente a $V(G)$. Formalmente, a centralidade de grau do tempo t_i é determinada por:

$$TDC(t_i) = \sum_{u \in V} deg((u, t_i))$$

Quanto maior for o valor da centralidade de grau do tempo de um instante de tempo, maior a importância desse instante de tempo em comparação aos demais instantes de tempo de um GVT.

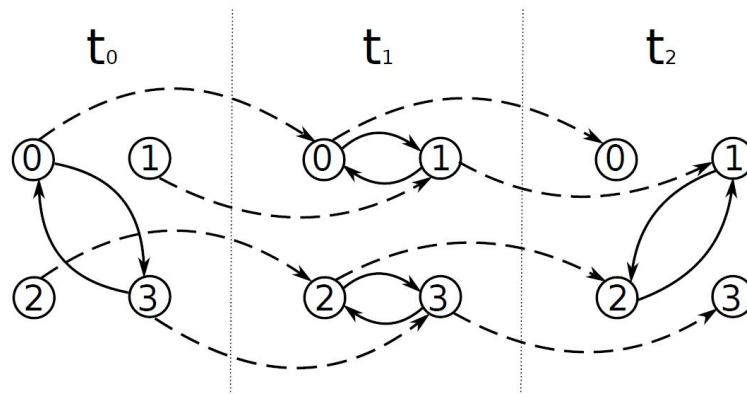


Figura 7 – Representação estrutural do GVT W

Considere como exemplo o GVT W mostrado na figura 7. Sejam o conjunto de nós $V(W) = \{0, 1, 2, 3\}$ e o conjunto de instantes de tempo $T(W) = \{t_0, t_1, t_2\}$. Para cada instante de tempo calcula-se o grau dos nós temporais. Dessa forma, para $t_i = t_0$ tem-se:

$$\begin{aligned} deg((0, t_0)) &= deg^+((0, t_0)) + deg^-((0, t_0)) = 2 + 1 = 3 \\ deg((1, t_0)) &= deg^+((1, t_0)) + deg^-((1, t_0)) = 1 + 0 = 1 \\ deg((2, t_0)) &= deg^+((2, t_0)) + deg^-((2, t_0)) = 1 + 0 = 1 \\ deg((3, t_0)) &= deg^+((3, t_0)) + deg^-((3, t_0)) = 2 + 1 = 3 \end{aligned}$$

Consequentemente, a centralidade de grau do tempo t_0 é:

$$TDC(t_0) = \sum_{u \in V} deg((u, t_0)) = 3 + 1 + 1 + 3 = 8$$

Analogamente, para $t_i = t_1$:

$$\begin{aligned} deg((0, t_1)) &= deg^+((0, t_1)) + deg^-((0, t_1)) = 2 + 2 = 4 \\ deg((1, t_1)) &= deg^+((1, t_1)) + deg^-((1, t_1)) = 2 + 2 = 4 \\ deg((2, t_1)) &= deg^+((2, t_1)) + deg^-((2, t_1)) = 2 + 2 = 4 \\ deg((3, t_1)) &= deg^+((3, t_1)) + deg^-((3, t_1)) = 2 + 2 = 4 \end{aligned}$$

Então, a centralidade de grau do tempo t_1 é dada por:

$$TDC(t_1) = \sum_{u \in V} deg((u, t_1)) = 16$$

Finalmente, para $t_i = t_2$:

$$\begin{aligned} deg((0, t_2)) &= deg^+((0, t_2)) + deg^-((0, t_2)) = 0 + 1 = 1 \\ deg((1, t_2)) &= deg^+((1, t_2)) + deg^-((1, t_2)) = 1 + 2 = 3 \\ deg((2, t_2)) &= deg^+((2, t_2)) + deg^-((2, t_2)) = 1 + 2 = 3 \\ deg((3, t_2)) &= deg^+((3, t_2)) + deg^-((3, t_2)) = 0 + 1 = 1 \end{aligned}$$

Dessa forma, a centralidade de grau do tempo t_2 é:

$$TDC(t_2) = \sum_{u \in V} deg((u, t_2)) = 8$$

Logo, conclui-se que o instante de tempo t_1 é o mais central dentre todos os instantes de tempo do GVT W , de acordo com a métrica de grau do tempo. Evidentemente, este é um exemplo simples que não considera o efeito de borda nos instantes de tempo t_0 e t_2 . Em redes maiores, com mais instantes de tempo, o efeito de borda é praticamente eliminado.

De modo complementar, foi realizado um estudo sobre a variação do grau do tempo, que simplesmente equivale à derivada discreta do grau do tempo. Sua finalidade é mostrar em quais instantes de tempo a métrica de grau do tempo variou mais abruptamente com relação à métrica de grau do tempo do instante de tempo imediatamente anterior.

Seja um instante de tempo $t_i \in T(G)$. A variação do grau do tempo t_i , $TDV(t_i)$, é dada por:

$$TDV(t_i) = \frac{TDC(t_i) - TDC(t_{i-1})}{TDC(t_{i-1})}, \text{ para } i \geq 1$$

O propósito é analisar a mudança do grau do instante de tempo com relação ao grau do instante de tempo imediatamente anterior, de forma a mostrar em quais instantes o grau do tempo variou mais significativamente.

A variação do grau do tempo pode ser positiva ou negativa, dependendo se o grau do tempo de um instante de tempo aumenta ou diminui, respectivamente, com relação ao grau do instante de tempo anterior. Quanto mais positiva for a variação do grau de um instante de tempo, mais importante ele é frente aos demais instantes de tempo.

Como exemplo, considere novamente o GVT W mostrado na figura 7. Uma vez que a variação é calculada a partir do segundo instante de tempo, para $t_i = t_1$, tem-se:

$$TDV(t_1) = \frac{TDC(t_1) - TDC(t_0)}{TDC(t_0)} = 1$$

Do mesmo modo, para $t_i = t_2$:

$$TDV(t_2) = \frac{TDC(t_2) - TDC(t_1)}{TDC(t_1)} = -\frac{1}{2}$$

Assim, conclui-se que o instante de tempo t_1 é o mais central dos instantes de tempo do GVT W , de acordo com a variação do grau do tempo.

5.1.2 Tempo de Cobertura

Sucintamente, a métrica de tempo de cobertura avalia o tempo necessário para a difusão de uma informação atingir um percentual ou um número dado de nós em um GVT, a partir de um determinado instante de tempo. Note que o conceito de nós atingidos se restringe ao conjunto de nós $V(G)$, ou seja, nós da forma (u, \bullet) . Nesse sentido, não importa em qual instante de tempo o nó u é atingido pela difusão, mas se ele é atingido ou não.

A difusão da informação segue um algoritmo de Busca em Largura (*Breadth-First Search*): um nó temporal genérico (u, t_i) inicia a difusão ao transmitir a informação para todos seus nós adjacentes. Por sua vez, os nós temporais adjacentes retransmitem a informação para seus próprios nós adjacentes no instante de tempo devido, e assim a difusão é realizada. A informação é difundida até que a quantidade (porcentagem) de nós τ seja atingida e o tempo necessário para a difusão é armazenado. Depois que este procedimento é repetido para todo nó $(u, t_i) \in VT(G)$, a média dos tempos de difusão é obtida. O valor da métrica de tempo de cobertura de t_i é justamente o tempo médio da difusão a partir de todos os nós que estão no instante de tempo t_i .

Por exemplo, a figura 8 mostra um processo de difusão iniciado em t_0 e terminado em t_{10} e outro processo iniciado em t_1 e terminado em t_9 . Intuitivamente, o instante de tempo t_1 é mais importante, uma vez que a difusão iniciada nele é mais rápida.

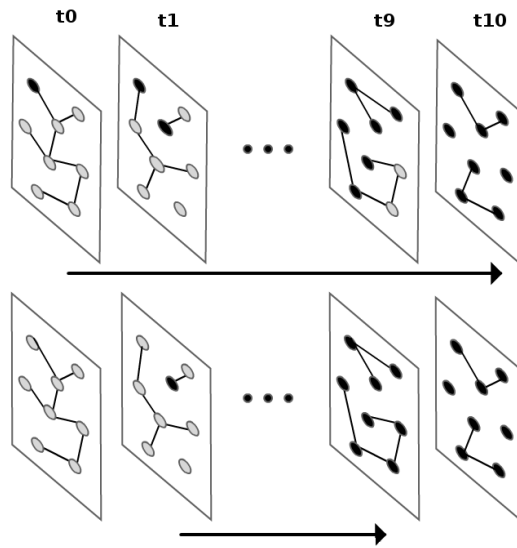


Figura 8 – Comparando diferentes processos de difusão

Considere $t_i \in T(G)$ o instante de tempo a ser avaliado pela métrica de tempo de cobertura. Seja $\mathcal{C}(t_i) \subseteq VT(G)$ o subconjunto de nós temporais do GVT G presentes na rede no instante de tempo t_i . Formalmente:

$$\mathcal{C}(t_i) = \{\forall(u, t) \in VT(G) | t = t_i\}$$

Sejam também τ a quantidade (ou a porcentagem) limite de nós a serem atingidos pela informação difundida; $AL((u, t))$ o conjunto de nós adjacentes a um nó temporal (u, t) . É importante frisar que dois nós temporais (u, t_i) e (v, t_j) são adjacentes se, e somente se, existe uma aresta dinâmica $e \in E(G)$ tal que $(u, t_i) = (\pi_1(e), \pi_2(e))$ e $(v, t_j) = (\pi_3(e), \pi_4(e))$. Considere também RL o conjunto dos nós atingidos pela informação difundida e WL o conjunto dos nós que ainda não receberam a informação difundida, mas que têm nós adjacentes que já a receberam. Também é necessário guardar em WL em qual instante de tempo os nós vão receber a informação (isso é conhecido de antemão, já que se trata de nós temporais). As entradas referente a esses nós permanecem em WL até o instante de tempo no qual eles se tornam aptos a retransmitir a informação.

Mais precisamente, o tempo de cobertura de t_i , $TC(t_i) = t_j - t_i$, onde t_j é o instante de tempo no qual a difusão atinge $\tau\%$ dos nós $\in V(G)$. Um nó (n, t_x) recebe a informação se, em qualquer instante de tempo entre t_x e t_i , existe uma aresta conectando n e um nó w , onde $w \in RL$. Inicialmente, RL é um conjunto unitário, que conta apenas com o nó que inicia a difusão.

Considere uma situação hipotética, com duas arestas dinâmicas $e_1 = (u, t_a, v, t_d)$ e $e_2 = (z, t_b, v, t_c)$, onde $u, v, z \in V(G)$; $t_a, t_b, t_c, t_d \in T(G)$ e $t_a < t_b < t_c < t_d$. Ao realizar-se a difusão, a transmissão a partir do nó temporal (u, t_a) será mais lenta do que

a transmissão a partir do nó temporal (z, t_b) , mesmo aquela tendo começado antes desta ($t_c < t_d$). Por isso, é importante manter sempre atualizado, em WL , o menor instante de tempo no qual o nó receptor w terminará de receber a informação difundida.

Observe que a difusão da informação não precisa necessariamente seguir uma ordem temporal, como ocorre no caso de arestas regressivas. Em contrapartida, o tempo de difusão é um aspecto estritamente positivo. Por exemplo: considere uma aresta mista $e = (u, t_3, v, t_1)$. Mesmo que o instante de tempo de origem seja maior que o instante de tempo de destino, a diferença entre eles é, de fato, positiva, e equivale a dois instantes de tempo.

A descrição completa do cálculo da métrica de tempo de cobertura de um instante de tempo t_i pode ser representada pelo algoritmo 1. O algoritmo deve ser aplicado a todos os instantes de tempo pertencentes a $T(G)$.

A complexidade do algoritmo é determinada basicamente pelos cinco laços que o compõem. A complexidade do laço mais interno (laço 5) é $O(n)$. Uma vez que o laço aninhado a ele (4) é executado n vezes, a complexidade dessa estrutura é igual a $O(n) \times O(n) = O(n^2)$. O laço anterior a essa estrutura (3) tem complexidade $O(n)$. Como existe um outro laço (2), executado n vezes, envolvendo o laço 3 e os laços aninhados 4 e 5, sua complexidade é dada por $O(n) \times (O(n) + (O(n^2))) = O(n^3)$. Visto que o laço mais externo (laço 1) também é executado n vezes, a complexidade do algoritmo completo é igual $O(n) \times O(n^3) = O(n^4)$.

A métrica de Tempo de Cobertura é inversamente proporcional à importância dos instantes de tempo: quanto menor for o tempo médio de cobertura a partir de um determinado instante de tempo, maior a importância desse instante de tempo em comparação aos demais instantes de tempo de um GVT. Isso significa que a difusão iniciada no instante de tempo melhor avaliado tende a ser mais eficiente, porque durou menos tempo do que a difusão nos demais instantes de tempo.

A fim de ilustrar o funcionamento do algoritmo que calcula a centralidade de tempo de cobertura, considere o GVT N representado na figura 9. Sejam o conjunto de nós $V(N) = \{0, 1, 2, 3, 4, 5, 6, 7\}$ e o conjunto de instantes de tempo $T(N) = \{t_0, t_1, t_2, t_3, t_4\}$. O conjunto de arestas dinâmicas ordenadas $E(N)$ é representado pelas ligações entre os nós. As arestas espaciais são representadas pelas linhas sólidas e as arestas mistas pelas linhas tracejadas. As arestas temporais estão representadas implicitamente no sistema. As arestas espaciais são, na verdade, determinadas por arestas orientadas e suas respectivas recíprocas. A porcentagem limite de nós a serem atingidos pela difusão τ é igual a 57% (equivalente a 4 nós). Por se tratar de um GVT relativamente pequeno (apenas 7 nós e 5 instantes de tempo), para o cálculo do tempo de cobertura serão considerados apenas os três primeiros instantes de tempo, t_0, t_1 e t_2 . Difusões iniciadas em t_3 e t_4 não são capazes de alcançar a quantidade de nós definida por τ .

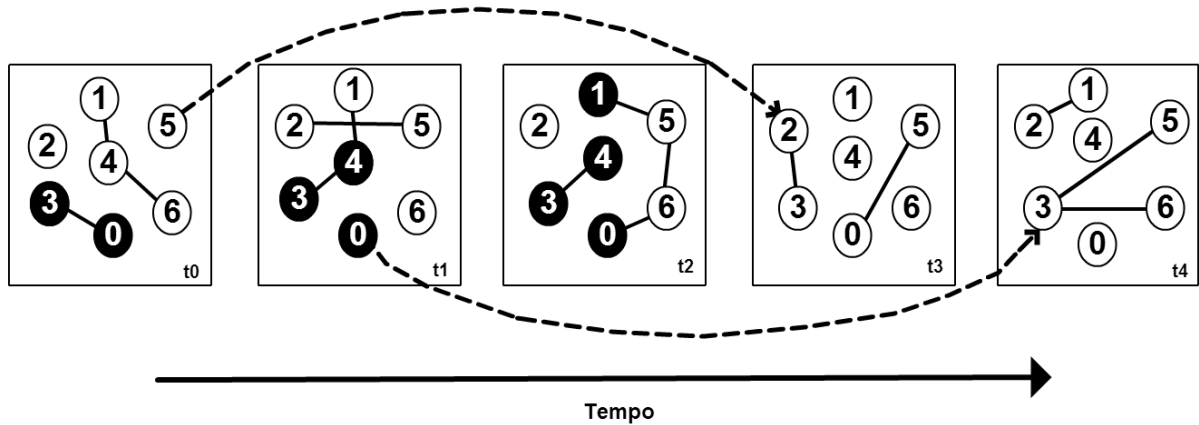


Figura 9 – Representação estrutural do GVT N e difusão iniciada no nó $(0, t_0)$

Para $t = t_0$, deve-se realizar uma difusão a partir de um nó temporal qualquer. Iniciando do nó $(0, t_0)$, a difusão alcança o nó 3 ainda em t_0 , o nó 4 em t_1 e o nó 1 em t_2 . Dessa forma, são alcançados 4 nós (incluindo o nó inicial) em 2 instantes de tempo, (exemplo da figura 9).

Analogamente, iniciando a partir do nó $(1, t_0)$, a difusão atinge 4 nós em 2 instantes de tempo. Para todos os demais nós em t_0 , a difusão atinge 4 nós também em 2 instantes de tempo. O tempo de cobertura para o instante de tempo inicial t_0 é dado pela média dos tempos de difusão a partir de todos os nós em t_0 . Logo, o tempo médio de cobertura do instante de tempo t_0 é igual 2 instantes de tempo.

Para $t = t_1$, os tempos de difusão a partir dos nós temporais $(0, t_1)$, $(1, t_1)$, $(2, t_1)$, $(3, t_1)$, $(4, t_1)$, $(5, t_1)$ e $(6, t_1)$ são 1, 1, 1, 3, 1, 1 e 3 instantes de tempo, respectivamente. Logo, o tempo médio de cobertura do instante t_1 é igual a 1,57 instantes de tempo.

Para $t = t_2$, os tempos de difusão a partir dos nós temporais $(0, t_2)$, $(1, t_2)$, $(2, t_2)$, $(3, t_2)$, $(4, t_2)$, $(5, t_2)$ e $(6, t_2)$ são 2, ∞ , 2, 2, 2, 1 e 2 instantes de tempo, respectivamente. A difusão iniciada no nó $(1, t_2)$ não atingiu os 4 nós necessários ($\tau = 57\%$) e, para efeito de simplificação, não é considerada no cálculo da média. Logo, o tempo médio de cobertura do instante de tempo t_1 é igual a 1,83 instantes de tempo.

Dessa forma, o menor tempo de cobertura encontrado é aquele cuja difusão foi iniciada no instante de tempo t_1 .

5.1.3 Centralidade de Cobertura

De maneira similar à métrica de tempo de cobertura, a métrica de centralidade de cobertura também utiliza a difusão de uma informação para avaliar os instantes de tempo. A diferença está justamente na escolha do aspecto a ser avaliado: enquanto o tempo de cobertura analisa o tempo necessário para a difusão atingir um percentual predeterminado

Algorithm 1 Cálculo da métrica de tempo de cobertura de um instante de tempo t_i

INPUT: *Time instant* t_i

OUTPUT: $\text{mean}(\text{time} \forall \text{node}(u, t_i))$

for all $\text{node}(u, t_i) \in VT(G)$ **do**

$\text{time} \leftarrow t_i$

Insert (u, \bullet) **into** RL

while $\text{size}(RL) < \tau$ **do**

for all $\text{node}(w, t_j) \in WL$ **do**

if $|t_j - \text{time}| = 0$ **then**

Remove all (w, \bullet) **from** WL

Insert (w, \bullet) **into** RL

end if

end for

for all $\text{node}(p, \bullet) \in RL$ **do**

if $\exists \text{node}(p, \text{time})$ **then**

for all $\text{node}(w, t_j) \in AL(p, \text{time})$ **do**

if $(w, \bullet) \notin RL$ **then**

(p, time) sends message to (w, t_j)

Insert (w, t_j) **into** WL

else

(p, time) sends message to (w, t_j)

end if

end for

end if

end for

$\text{time} \leftarrow \text{time} + 1$

end while

end for

de nós de um GVT, a centralidade de cobertura estima a quantidade (ou percentual) de nós de um GVT que é atingida pela difusão da informação, dado um número de passos fixo φ do algoritmo de busca em largura.

A essência do algoritmo de busca em largura é justamente o processamento dos nós em níveis: a princípio, são descobertos os nós que estão a $d = 1$ passos do nó que inicia uma busca (difusão); posteriormente, são descobertos os nós a $d = 2$ passos, $d = 3$ passos, e assim sucessivamente. No caso da centralidade de cobertura, o número máximo de passos permitido é $d = \varphi$.

Por exemplo, a figura 10 apresenta dois diferentes processos de difusão. Ambos os processos ocorrem em φ passos. Percebe-se que o processo que começa em t_5 atinge um maior número de nós. Conseqüentemente, define-se que t_5 é mais importante que t_0 .

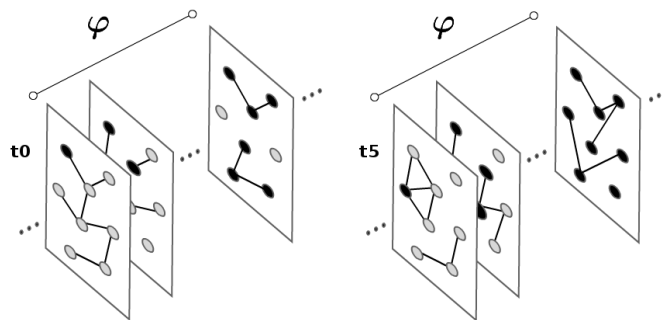


Figura 10 – Comparando diferentes processos de difusão

Note, mais uma vez, que o conceito de nós atingidos se restringe ao conjunto de nós $V(G)$, que são nós da forma (u, \bullet) . Ou seja, nesse contexto não importa em qual instante de tempo o nó u é atingido pela difusão, mas se ele foi atingido ou não.

Existem algumas estruturas auxiliares para coordenar a execução do procedimento que calcula a métrica de centralidade de cobertura. Normalmente, utiliza-se um esquema com três cores para diferenciação do estado de cada nó: *WHITE* (nó ainda não atingido pela difusão), *GREY* (nó atingido, mas seus nós adjacentes ainda não foram atingidos) e *BLACK* (nó atingido, com seus nós adjacentes também atingidos). Além disso, sejam $color[(u, t)]$ a estrutura que mantém a informação sobre o estado do nó (u, t) ; $d[(u, t)]$ a estrutura que mantém a quantidade de passos entre o nó que inicia a difusão e o nó (u, t) ; RL o conjunto dos nós que foram atingidos pela difusão da informação e $AL(u, t)$ o conjunto de nós adjacentes ao nó (u, t) .

Mais precisamente, assume-se um processo de difusão semelhante ao definido para o tempo de cobertura. Além disso, a centralidade de cobertura de t_i é $CC(t_i) = b$, onde $b = [0, 100]$ e b representa a porcentagem total de rede alcançada por uma difusão dentro de uma restrição de passos.

O procedimento descrito retorna o tamanho de RL , que é efetivamente a quantidade

(percentual) de nós atingidos pela difusão. Depois que ele é repetido para todo nó $u \in \mathcal{C}(t_i)$, a média da quantidade de nós atingidos pela difusão é obtida. O valor da métrica de centralidade de cobertura de t_i é justamente a quantidade média de nós atingidos pela difusão a partir de todos os nós que estão no instante de tempo t_i .

O cálculo da métrica de centralidade de cobertura de um instante de tempo t_i pode ser sintetizada pelo algoritmo 2. O algoritmo deve ser aplicado para todos os instantes de tempo pertencentes a $T(G)$.

A complexidade do algoritmo é determinada basicamente pelos quatro laços que o compõem. A complexidade do laço mais interno (laço 4) é $O(n)$. Uma vez que o laço aninhado a ele (3) é executado n vezes, a complexidade dessa estrutura é igual a $O(n) \times O(n) = O(n^2)$. O laço anterior a essa estrutura (2) tem complexidade $O(n)$. Como ainda existe um laço mais externo (1), executado n vezes, envolvendo o laço 2 e os laços aninhados 3 e 4, a complexidade do algoritmo é dada por $O(n) \times (O(n) + (O(n^2))) = O(n^3)$.

A métrica de Centralidade de cobertura é diretamente proporcional à importância dos instantes de tempo: quanto maior for a quantidade (percentual) média de nós atingidos pela difusão a partir de um determinado instante de tempo, maior a importância desse instante de tempo em comparação aos demais em um GVT. Pode-se dizer que a difusão iniciada no instante de tempo melhor avaliado tende a ser mais eficiente, pois atingiu, em média, um número maior de nós do que a difusão iniciada nos demais instantes de tempo.

A fim de ilustrar o funcionamento do algoritmo que calcula a centralidade de cobertura, considere mais uma vez o GVT N representado na figura 9. Novamente, serão desconsiderados os instantes de tempo t_3 e t_4 . Para esse exemplo, o número fixo de passos φ é igual a 2.

Para $t = t_0$, realiza-se a difusão a partir de cada nó temporal, independentemente. Iniciando do nó $(0, t_0)$, a difusão alcança o nó $(3, t_0)$ em 1 passo e o nó $(4, t_1)$ em 2 passos. Ao todo, a difusão atingiu 3 nós (incluindo o nó inicial). A partir do nó $(1, t_0)$, a difusão atinge 3 nós em 2 passos. As difusões iniciadas nos nós $(2, t_0)$, $(3, t_0)$, $(4, t_0)$, $(5, t_0)$ e $(6, t_0)$ atingem, em 2 passos, 2, 3, 4, 2 e 3 nós, respectivamente. Logo, a centralidade de cobertura média, iniciada no instante de tempo t_0 , é igual a 2,85 nós (40,7% da rede).

Para $t = t_1$, as difusões iniciadas nos nós $(0, t_1)$, $(1, t_1)$, $(2, t_1)$, $(3, t_1)$, $(4, t_1)$, $(5, t_1)$ e $(6, t_1)$ atingem, em 2 passos, 4, 4, 4, 2, 4, 4 e 3 nós, respectivamente. Logo, a centralidade de cobertura média, iniciada no instante de tempo t_1 , é igual a 3,57 nós (51% da rede).

Para $t = t_2$, as difusões iniciadas nos nós $(0, t_2)$, $(1, t_2)$, $(2, t_2)$, $(3, t_2)$, $(4, t_2)$, $(5, t_2)$ e $(6, t_2)$ atingem, em 2 passos, 3, 3, 2, 3, 2, 4 e 3 nós, respectivamente. Logo, a centralidade de cobertura média, iniciada no instante de tempo t_2 , é igual a 3 nós (42,8% da rede).

Dessa forma, a maior centralidade de cobertura encontrada foi aquela cuja difusão foi iniciada no instante de tempo t_1 .

Algorithm 2 Cálculo da métrica de centralidade de cobertura de um instante de tempo t_i

INPUT: *Time instant* t_i

OUTPUT: $\text{mean}(\text{size}(RL) \forall \text{node}(v, t_i))$

```

for all  $\text{node}(v, t_i) \in VT(G)$  do
  for all  $\text{node}(u, t_j) \in VT(G) - (v, t_i)$  do
     $\text{color}[(u, t_j)] \leftarrow WHITE$ 
     $d[(u, t_j)] \leftarrow \infty$ 
  end for

   $\text{color}[(v, t_i)] \leftarrow GREY$ 
   $d[(v, t_i)] \leftarrow 0$ 
  Insert  $(v, t_i)$  into Queue
  Insert  $(v, \bullet)$  into RL

  while Queue  $\neq \emptyset$  do

    Remove  $\text{node}(u, t_j)$  from Queue

    for all  $\text{node}(p, t_k) \in AL((u, t_j))$  do

      if  $\text{color}[(p, t_k)] = WHITE$  then
         $(u, t_j)$  sends message to  $(p, t_k)$ 
         $\text{color}[(p, t_k)] \leftarrow GREY$ 
         $d[(p, t_k)] \leftarrow d[(u, t_j)] + 1$ 

        if  $d[(p, t_k)] > \varphi$  then
          store  $\text{size}(RL)$ 
        end if

        Insert  $(p, t_k)$  into Queue

        if  $(p, \bullet) \notin RL$  then
          Insert  $(p, \bullet)$  into RL
        end if

      end if

    end for

     $\text{color}[(u, t_j)] \leftarrow BLACK$ 

  end while

end for

```

end for

6 APLICAÇÃO E ANÁLISE DE CENTRALIDADE DE TEMPO

Neste capítulo é apresentada a avaliação dos resultados das métricas desenvolvidas, com base em conjunto de dados real, obtido de uma rede de contatos presenciais em um ambiente hospitalar. Na seção 6.1, o conjunto de dados é descrito. Na seção 6.2, são discutidos os resultados do Grau do Tempo e a influência do tamanho dos *snapshots* da rede sobre a métrica de Grau do Tempo. Na seção 6.3, é analisada a variação da métrica de Grau do Tempo entre instantes de tempo subsequentes. Na seção 6.4, são mostrados os resultados da métrica de Tempo de Cobertura e a relação entre os instantes de tempo melhor qualificados e outros escolhidos aleatoriamente. Na seção 6.5, são mostrados os resultados da métrica de Centralidade de Cobertura e a relação entre os instantes de tempo melhor qualificados e outros escolhidos aleatoriamente.

6.1 CONJUNTO DE DADOS MOSAR

O conjunto de dados utilizado neste trabalho foi disponibilizado pelo projeto MOSAR (*Mastering hOSpital Antimicrobial Resistance and its spread*) (LUCET et al., 2012). O MOSAR é um projeto de colaboração científica entre diversos institutos de pesquisa médica, bioquímica e em computação que visa ampliar o conhecimento na dinâmica de transmissão de bactérias resistentes a antimicrobianos (AMRB) em ambientes de alto risco, como unidades de terapia intensiva, centros cirúrgicos e centros de reabilitação de saúde em países europeus com altos níveis de AMRB.

Os dados foram coletados a partir de contatos presenciais entre pessoas (médicos, enfermeiros, funcionários e pacientes) em uma determinada ala de um hospital durante um período de duas semanas (entre a 0h do dia 25/07/2009 e a 0h do dia 08/08/2009). Cada uma das 160 pessoas que participaram do experimento foi equipada com um dispositivo RFID que detecta a presença de outros dispositivos em um raio de abrangência pequeno (menor ou igual a um metro). A identificação de cada dispositivo é única e está associada sempre a uma mesma pessoa. A cada 30 segundos, cada dispositivo verifica e armazena a lista de todos os dispositivos (nós) que se encontram dentro de seu raio de abrangência, de forma a estabelecer a disposição dos contatos entre eles (arestas).

Pode-se modelar o sistema descrito anteriormente como um GVT $M = (V, E, T)$, onde $V(M)$ é o conjunto de todas as pessoas que passam pela ala do hospital e estão munidas com o dispositivo RFID, $E(M)$ é o conjunto de todos os contatos presenciais estabelecidos entre as pessoas e $T(M)$ é o conjunto sequencial de instantes de tempo com intervalo de 30 segundos entre si. Dessa forma, $|V(M)| = 160$ e $|T(H)| = 40320$.

O contato entre duas pessoas é claramente representado por uma aresta não orientada. Como as arestas no modelo unificado são descritas por quádruplas ordenadas, há a necessidade de que a ligação entre quaisquer dois nós temporais (u, t_a) e (v, t_b) seja

representada pela aresta (u, t_a, v, t_b) e sua recíproca (v, t_b, u, t_a) , simultaneamente.

Evidentemente, não é possível que, em uma rede de contatos presenciais, uma pessoa estabeleça uma conexão consigo mesma em um mesmo instante de tempo e nem que duas pessoas diferentes em instantes de tempo diferentes estejam conectadas. Sendo assim, as classes de arestas que não são encontradas no GVT M são as arestas dinâmicas mistas $((u, t_a, v, t_b)$, com $t_a \neq t_b$) e as arestas *self-loop* $((u, t_a, u, t_a))$.

O GVT M é naturalmente construído pelas demais classes de arestas dinâmicas. As arestas espaciais $((u, t_a, v, t_a))$, com $u \neq v$) determinam a disposição dos contatos entre diferentes pessoas em um mesmo instante de tempo (*snapshot*). O uso de arestas temporais $((u, t_a, u, t_b)$, com $t_a \neq t_b$) pode ser justificado pelo fato de que elas representam um estado de espera de um nó. Uma pessoa que carrega consigo um agente patológico em um instante de tempo t_a tende a mantê-lo em um instante de tempo t_b , com $t_a < t_b$, dependendo do modelo epidêmico em questão.

6.2 GRAU DO TEMPO

O primeiro resultado a ser analisado é o Grau do Tempo, com o tamanho padrão do *snapshot* igual a 30 segundos. Neste caso, tamanho do *snapshot* se refere ao intervalo de tempo entre *snapshots* que representam a estrutura da rede em instantes de tempo subsequentes.

A figura 11 mostra a Função de Distribuição Cumulativa (CDF) da métrica de Grau do Tempo. De acordo com os resultados, o Grau do Tempo médio dos instantes de tempo da rede MOSAR é igual a 41,6, com Grau do Tempo máximo e mínimo igual a 74 e 12, respectivamente. Ao analisar a relação entre os 5% melhor qualificados instantes de tempo pela métrica de Grau do Tempo e os demais instantes de tempo, chegamos a um valor médio igual a 66,8 para os melhores e 36,4 para os demais. Isso mostra uma diferença de aproximadamente 85% entre o Grau do Tempo médio dos dois grupos.

A próxima análise a ser feita é a relação entre a centralidade de grau do tempo e o tamanho do *snapshot*. As figuras 12, 13, 14 e 15 mostram a CDF da métrica de grau do tempo para diferentes tamanhos de *snapshot*: 30 segundos, 1 minuto, 2 minutos e 4 minutos, respectivamente. Apesar do aumento gradativo no tamanho dos *snapshots*, é possível observar que a distribuição da métrica de grau do tempo tem comportamento similar em todos os casos, apesar de estar deslocada.

Com o aumento no tamanho dos *snapshots*, também é perceptível o aumento da centralidade de grau do tempo entre os instantes de tempo. Isso é esperado, uma vez que *snapshots* maiores agregam mais arestas entre os nós temporais, o que impacta diretamente o cálculo da métrica de grau do tempo. Por exemplo, o grau médio do tempo é 41,6 para o tamanho de *snapshot* igual a 30 segundos. Para os demais tamanhos de *snapshot*

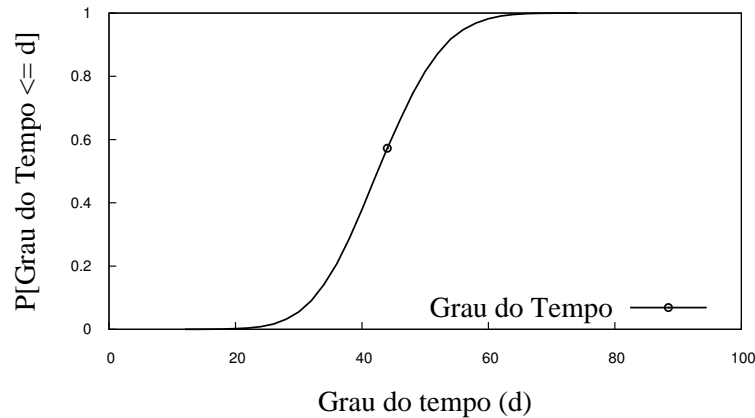


Figura 11 – Função de distribuição cumulativa do Grau do Tempo

(1 minuto, 2 minutos e 4 minutos), o grau médio do tempo é igual a 48,1, 53,1 e 61,1, respectivamente.

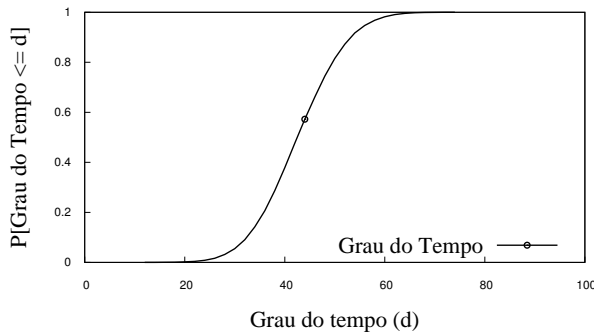
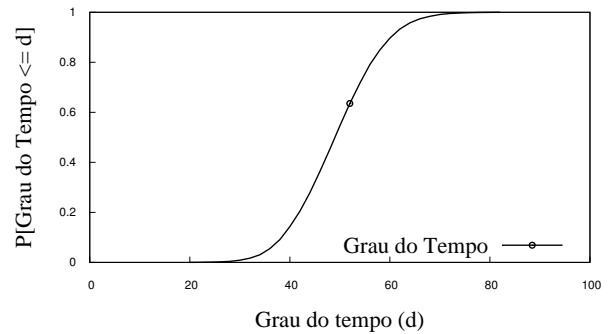
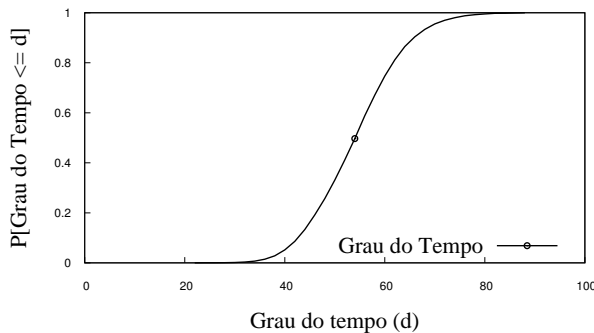
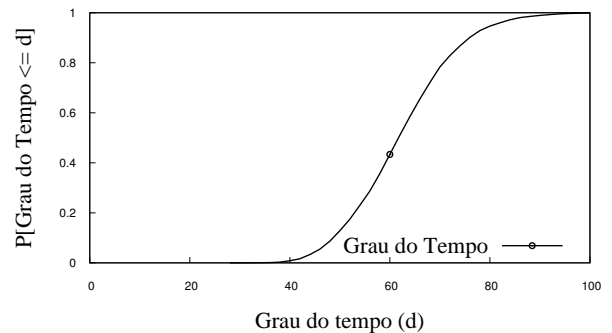
Outra análise a ser considerada é se instantes de tempo mais importantes implicariam ou não na existência de outros instantes de tempo mais importantes, quando o tamanho do *snapshot* varia. Para isso, comparamos os 5% melhores instantes de tempo para os diferentes tamanhos de *snapshots* analisados, utilizando o coeficiente de *Jaccard* entre cada par de classificações. O coeficiente de *Jaccard* é definido por $J = C_1 \cap C_2 / C_1 \cup C_2$, onde C_1 e C_2 são os conjuntos a serem comparados. O coeficiente pode variar de 0 (nenhuma similaridade) a 1 (similaridade total).

Foi observado que há uma similaridade relativamente alta entre as classificações dos 5% melhores instantes de tempo, segundo a métrica de Grau do Tempo, nos quatro cenários analisados proporcionalmente. A similaridade entre as classificações para tamanhos de *snapshots* de 30 segundos e 1 minuto é igual a 0.67; entre 1 minuto e 2 minutos, igual a 0.65; entre 2 minutos e 4 minutos, igual a 0.66.

Dessa forma, pode-se concluir que o tamanho do *snapshot* não influencia tão intensamente a classificação dos instantes de tempo, uma vez que há uma tendência dos instantes de tempo melhor qualificados implicarem na existência de instantes de tempo importantes nas demais variações da rede.

6.3 VARIAÇÃO DO GRAU DO TEMPO

Também foi analisada a variação (derivada discreta) do grau do tempo da rede original do MOSAR (tamanho do *snapshot* igual a 30 segundos). Nesse caso, o objetivo

Figura 12 – Tamanho do *snapshot*: 30 segundosFigura 13 – Tamanho do *snapshot*: 1 minutoFigura 14 – Tamanho do *snapshot*: 2 minutosFigura 15 – Tamanho do *snapshot*: 4 minutos

é analisar a mudança do grau do instante de tempo com relação ao grau do instante de tempo imediatamente anterior, a fim de quantificar a variação do grau do tempo.

A figura 16 mostra a CDF da variação do grau do tempo. É possível observar que 90% das variações estão entre -0.2 e 0.2, o que indica que a maioria dos graus dos instantes de tempo não varia significativamente com relação ao instante de tempo anterior.

Uma variação positiva, com valores entre 0.2 e 1.25, e uma variação negativa, entre -0.6 e -0.2, podem ser observadas em cerca de 10% das amostras. Isso pode ser justificado por uma movimentação intensa e pontual na ala do hospital, normalmente decorrente de um evento externo, como chegada de ambulâncias ou trocas de turnos.

6.4 TEMPO DE COBERTURA

Como dito anteriormente, a métrica de Tempo de Cobertura avalia o tempo necessário (ou seja, a quantidade de instantes de tempo) para a difusão de uma informação

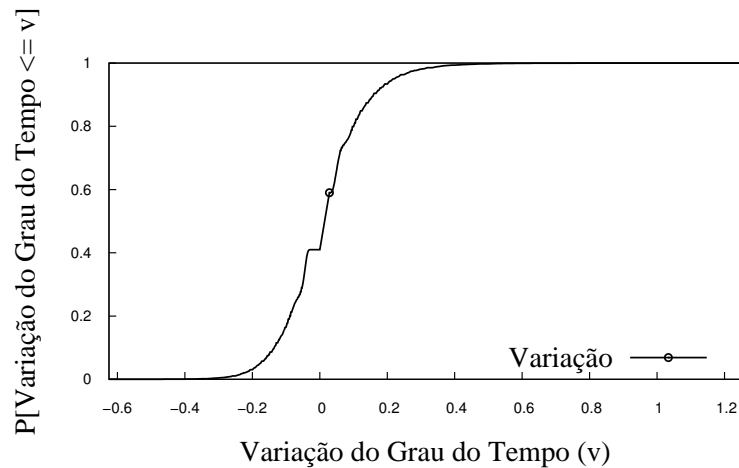


Figura 16 – Função de distribuição cumulativa da Variação do Grau do Tempo

atingir um percentual ou um dado número de nós τ em um GVT, a partir de um instante de tempo t_i . Deve-se salientar que os nós atingidos nesse caso se referem ao conjunto de nós $V(M)$. Intuitivamente, menores valores de Tempo de Cobertura para instantes de tempo indicam que estes podem difundir a informação mais rapidamente.

Neste trabalho, nós limitamos a 60% da rede (96 pessoas) o número a ser alcançado pela difusão. Avaliamos a rede e verificamos que esse número é o limite possível para que a difusão da informação pudesse ser realizada sem prejuízo de avaliação para a maioria dos instantes de tempo. Vários fatores podem contribuir para isso, como a variação significativa da quantidade de pessoas ao longo do tempo e a não recorrência de pessoas no sistema (por exemplo, pessoas que aparecem em instantes de tempo pontuais e não retornam mais ao hospital).

Da mesma forma, limitamos aos primeiros 37000 instantes de tempo do conjunto $T(M)$ a quantidade de instantes a serem avaliados pela métrica. Esse é o número máximo possível, pois difusões iniciadas após $t = 37000$ podem não atingir a quantidade máxima de 60% da rede estabelecida, levando a análises distorcidas da métrica de Tempo de Cobertura.

A figura 17 mostra a Função de Distribuição Cumulativa Complementar (CCDF) do Tempo de Cobertura para a rede MOSAR, variando a porcentagem da rede a ser atingida pela difusão. Neste trabalho, são considerados seis cenários diferentes: 10, 20, 30, 40, 50 e 60%. Foi utilizada a CCDF para uma melhor visualização dos resultados, pois o tempo de cobertura é inversamente proporcional à importância do instante de tempo.

De acordo com os resultados, para atingir 10% da rede MOSAR, o tempo de cobertura médio é de cerca de 646 instantes de tempo. Como o tamanho dos *snapshots*

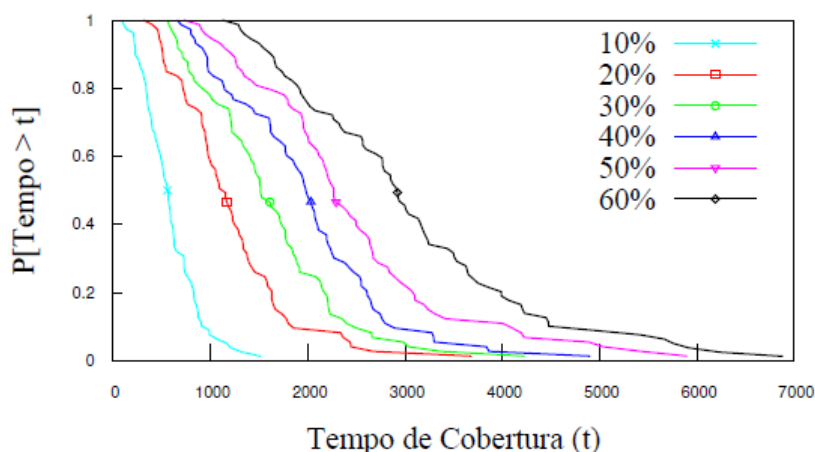


Figura 17 – Função de distribuição cumulativa do Tempo de Cobertura

da rede é de 30 segundos, pode-se dizer que a difusão demora 323 minutos (5 horas e 23 minutos) para atingir 10% da rede. De maneira similar, para atingir 20% da rede, o tempo de cobertura médio é de 1302 instantes de tempo; para 30%, 1718 instantes de tempo; para 40%, 2123; para 50%, 2624; e para atingir 60% da rede, foram necessários 3419 instantes de tempo (aproximadamente 28 horas e 30 minutos), em média.

Outra análise que pode ser feita é a relação entre os instantes de tempo melhor qualificados pela métrica de Tempo de Cobertura e os demais instantes de tempo. As figuras 18, 19, 20, 21, 22 e 23 mostram a CDF do Tempo de Cobertura para os 5% melhores instantes de tempo e para 5% dos instantes de tempo escolhidos aleatoriamente entre os demais em todos os cenários.

Segundo os resultados, o tempo de cobertura médio para atingir 10% da rede é de 266 instantes de tempo, com a difusão iniciada nos 5% melhores instantes de tempo, enquanto para os demais é de 680 instantes de tempo. Isso indica uma diferença de 155%, ou seja, o tempo de difusão para os instantes de tempo melhor qualificados é cerca de 2,5 vezes mais rápido.

Para atingir 20% da rede, o tempo médio de cobertura para os instantes de tempo melhor qualificados é igual a 535 instantes de tempo, contra 1370 dos demais. Isso aponta uma diferença de 156%, que representa uma difusão cerca de 2,5 vezes mais rápida em favor dos instantes melhor qualificados.

O comportamento se mantém para os demais cenários. Para atingir 30%, a difusão iniciada nos instantes de tempo melhor qualificados leva 729 instantes de tempo, o que equivale a uma diferença de 147% com relação aos demais instantes de tempo (aproximadamente 1806), cerca de 2,4 vezes mais rápida. Para a difusão atingir 40% da rede, foram necessários 976 instantes de tempo para os melhor qualificados e 2226 para os demais (diferença de 127%). Para atingir 50% da rede, a difusão demorou cerca de 1266

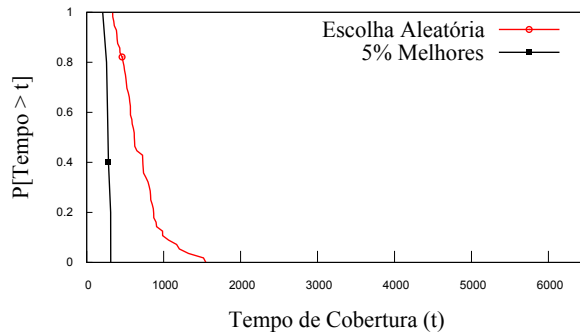


Figura 18 – Cobertura da rede: 10%

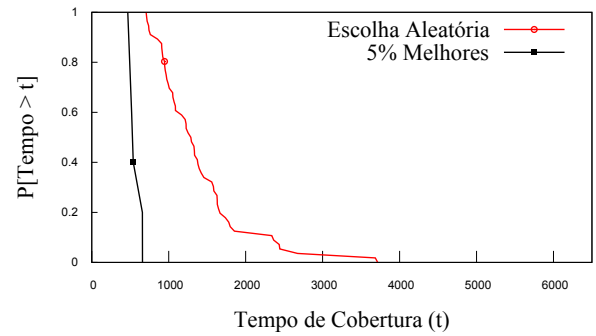


Figura 19 – Cobertura da rede: 20%

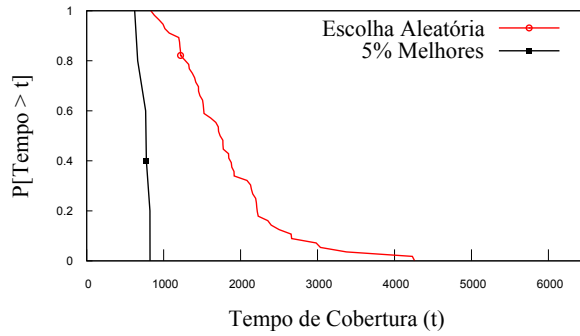


Figura 20 – Cobertura da rede: 30%

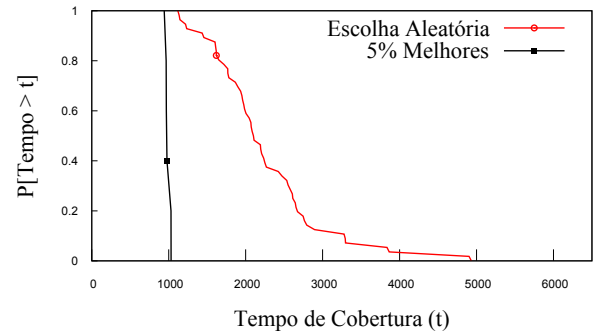


Figura 21 – Cobertura da rede: 40%

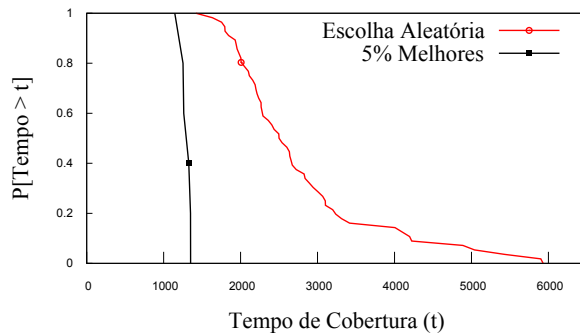


Figura 22 – Cobertura da rede: 50%

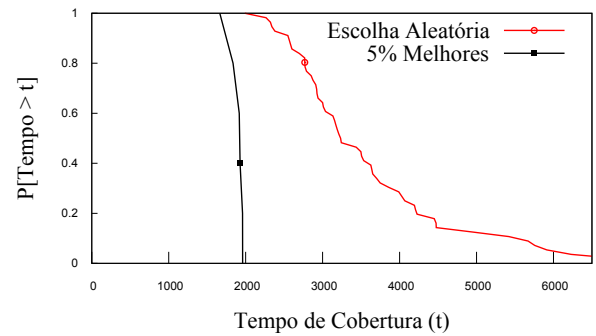


Figura 23 – Cobertura da rede: 60%

instantes de tempo para os melhor qualificados e 2745 para o demais (diferença de 116%).

Para atingir o tamanho máximo, limitado em 60% da rede, a difusão demorou, em

média, 1860 instantes de tempo para os 5% melhores instantes de tempo iniciais, uma diferença de 91% com relação aos demais instantes de tempo (cerca de 3558).

A diferença é ainda maior quando são comparados os 10 melhores instantes de tempo e os demais instantes de tempo. Como visto na figura 24, o tempo de cobertura médio para atingir 10% da rede é de 210 instantes de tempo, com a difusão iniciada nos 10 melhores instantes de tempo, enquanto para os demais é de 646 instantes de tempo. Isso indica uma diferença de 207%, ou seja, o tempo de difusão para os instantes de tempo melhor qualificados é cerca de 3,07 vezes mais rápido.

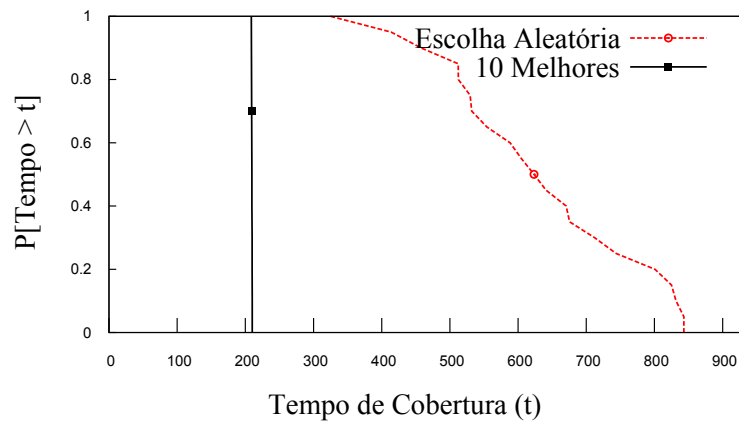


Figura 24 – Tempo de Cobertura (10% de Cobertura da rede): 10 Melhores × Demais instantes

Para finalizar a análise dos resultados da métrica de Tempo de Cobertura, é mostrada na figura 25 a matriz de similaridade entre as classificações dos 5% melhor qualificados instantes de tempo nos seis cenários estudados. Cada cruzamento de linha e coluna dessa matriz representa a similaridade, segundo o coeficiente de *Jaccard*, entre cada par de classificações. O coeficiente pode variar de 0 (nenhuma similaridade, cor branca) a 1 (similaridade total, cor azul-escuro).

Desconsiderando a diagonal principal, que mostra a similaridade de uma classificação com ela mesma, é possível ver que os maiores coeficientes de similaridade ocorrem entre as classificações dos cenários diretamente mais próximos, como 20 e 30%, 30 e 40% e 40 e 50% (0.82, 0.82 e 0.67, respectivamente). Isso é esperado, uma vez que os instantes de tempo melhor qualificados tendem a permanecer bem qualificados com o andamento do processo de difusão, e quaisquer mudanças que porventura acontecerem entre as classificações dos cenários mais próximos ocasionarão em menores impactos na similaridade. Por exemplo, a similaridade entre os cenários cujos tamanhos da rede a serem atingidos são 30 e 40% é de 0.82, enquanto para os cenários 10 e 60% é de apenas 0.11.

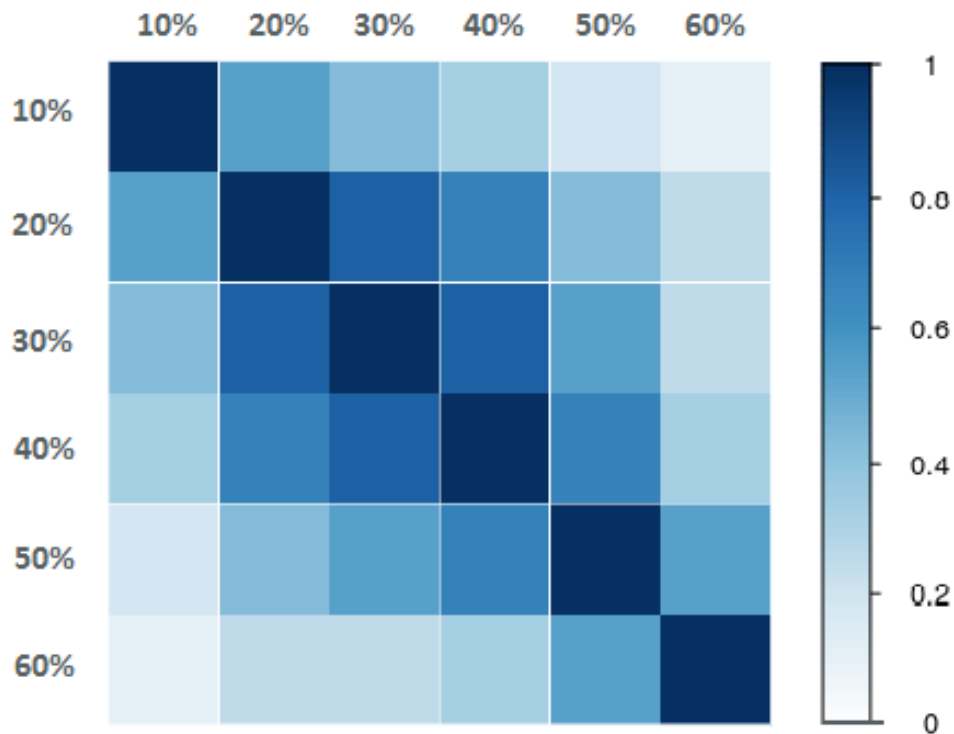


Figura 25 – Tempo de Cobertura: matriz de similaridade entre as classificações

6.5 CENTRALIDADE DE COBERTURA

Por definição, a centralidade de cobertura avalia a quantidade (ou percentual) de nós de um GVT que é atingida pela difusão de uma informação, dado um número fixo de passos φ . De fato, a distância entre o nó que inicia a difusão e todos os nós que são atingidos pela mesma não pode ultrapassar o número de passos fixo. Os nós atingidos também se referem ao conjunto $V(M)$.

Intuitivamente, maiores valores de Centralidade de Cobertura de instantes de tempo iniciais indicam que estes podem difundir a informação para uma número maior de nós.

Em razão dos mesmos obstáculos para o cálculo da métrica de tempo de cobertura, o tamanho máximo da rede a ser atingida pela difusão também equivale a 60% do nós de $V(M)$. De maneira análoga à métrica anterior, a análise da difusão a partir de instantes de tempo mais avançados foi descartada nos resultados finais.

A figura 26 mostra a CDF da Centralidade de Cobertura para a rede MOSAR, variando o número fixo de passos da difusão. Neste trabalho, são considerados quatro cenários diferentes: 500, 1000, 1500 e 2000 passos.

De acordo com os resultados, para uma difusão com 500 passos, a centralidade de cobertura média é de cerca de 10% da rede (16 pessoas). Analogamente, para uma difusão com 1000 passos, a fração da rede atingida é de aproximadamente 19% (31 pessoas); para

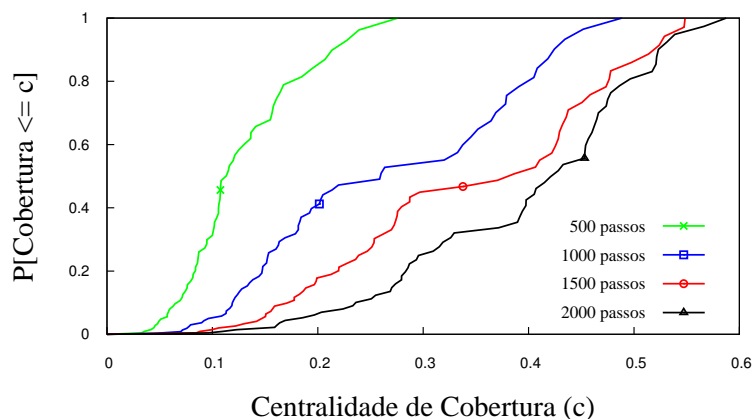


Figura 26 – Função de distribuição cumulativa da Centralidade de Cobertura

uma difusão com 1500 passos, o tamanho atingido da rede é de 27,5% (44 pessoas); para uma difusão com 2000 passos, a fração da rede atingida é de 34,7% (55 pessoas).

Além disso, também foi feita a relação entre os 5% melhores instantes de tempo qualificados pela métrica de Centralidade de Cobertura e 5% dos demais instantes de tempo escolhidos aleatoriamente. As figuras 27, 28, 29 e 30 mostram a CDF da Centralidade de Cobertura para todos os cenários analisados.

Segundo os resultados, a centralidade de cobertura média para uma difusão com 500 passos é de 18,7% para os instantes de tempo melhor qualificados, enquanto que, para os demais instantes de tempo é de cerca de 9,2%. Isso mostra uma diferença de 103%, ou seja, a difusão iniciada nos instantes de tempo melhor qualificados atinge mais que o dobro das pessoas que aquela começada em instantes de tempo escolhidos aleatoriamente.

Para uma difusão de 1000 passos, a centralidade de cobertura média é de 40,7% para os instantes de tempo melhor qualificados e de 17% para os demais instantes de tempo. A difusão iniciada nos instantes de tempo melhor qualificados atingiu uma fração da rede 139% maior que para os demais instantes de tempo (2,4 vezes mais).

Para uma difusão de 1500 passos, a centralidade de cobertura média é de 50,4% para os instantes de tempo melhor qualificados e de 25,4% para os demais instantes de tempo. A quantidade de pessoas atingidas pela difusão iniciada em instantes de tempo melhor qualificados é quase o dobro (98%) com relação à difusão iniciada nos instantes de tempo escolhidos aleatoriamente.

Finalmente, para uma difusão de 2000 passos, a centralidade de cobertura média é de 54,4% para os instantes de tempo melhor qualificados e de 33% para os demais instantes

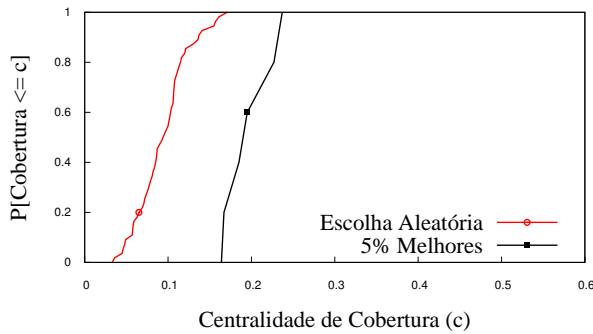


Figura 27 – Número fixo de passos: 500

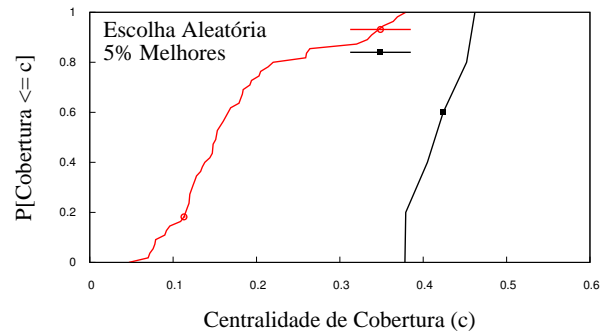


Figura 28 – Número fixo de passos: 1000

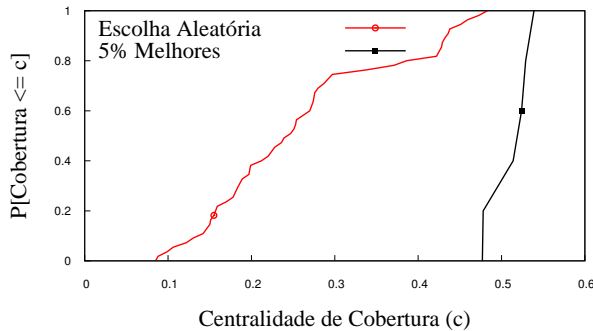


Figura 29 – Número fixo de passos: 1500

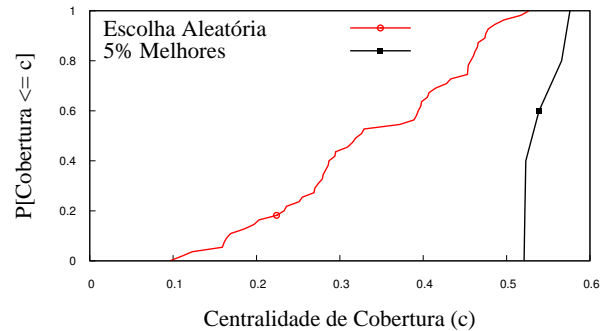


Figura 30 – Número fixo de passos: 2000

de tempo. A quantidade de pessoas atingidas pela difusão é 63% maior.

A diferença é ainda maior quando são comparados os 10 melhores instantes de tempo e os demais instantes de tempo. Por exemplo, como visto na figura 31, a centralidade de cobertura média para uma difusão de 2000 passos é de 56% , com a difusão iniciada nos 10 melhores instantes de tempo, enquanto para os demais é de 34%. Isso indica uma cobertura 64,7% maior.

Para concluir a análise dos resultados da métrica de Centralidade de Cobertura, é mostrada na figura 32 a matriz de similaridade entre as classificações dos 5% melhor qualificados instantes de tempo nos quatro cenários estudados. Mais uma vez foi utilizado o coeficiente de *Jaccard* para se chegar às similaridades entre as classificações. O coeficiente pode variar de 0 (nenhuma similaridade) a 1 (similaridade total).

Desta vez, é possível perceber que a similaridade entre as classificações não é tão alta, se comparada com os resultados da métrica de Tempo de Cobertura. Entretanto, nota-se que a similaridade não varia tanto como na métrica anterior (0.33 a 0.54 em

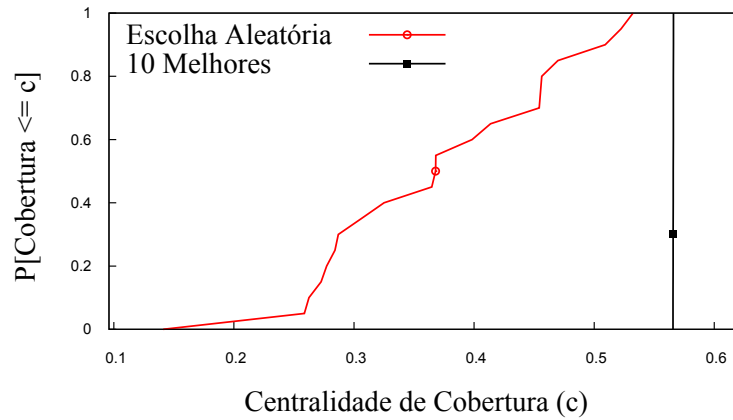


Figura 31 – Centralidade de Cobertura (2000 passos): 10 Melhores \times Demais instantes

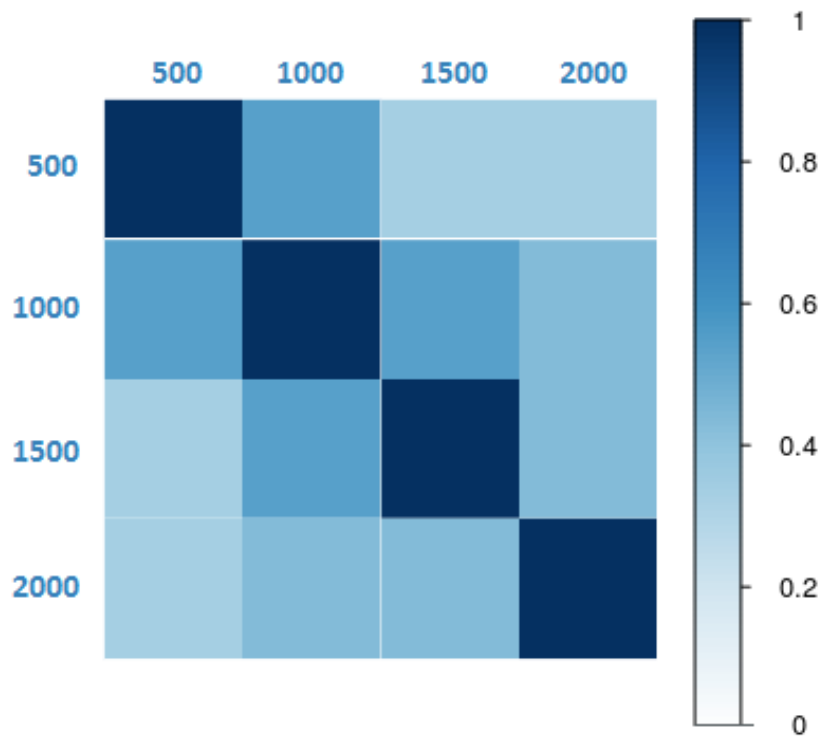


Figura 32 – Centralidade de Cobertura: matriz de similaridade entre as classificações

centralidade de cobertura contra 0.11 a 0.82 em tempo de cobertura). Mais uma vez, percebe-se que os maiores coeficientes de similaridade ocorrem entre as classificações dos cenários diretamente mais próximos (500 e 1000, 1000 e 1500, e 1500 e 2000 passos, respectivamente 0.54, 0.54 e 0.43).

7 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, definimos Centralidade de Tempo em GVTs. De maneira similar à centralidade de nó, a centralidade de tempo avalia a importância de um dado instante de tempo em um GVT. Para algumas aplicações modeladas em cenários de GVT, mais importante do que identificar os nós centrais sob uma determinada definição, é identificar os principais instantes de tempo para a tomada de decisões.

Até onde sabemos, este é o primeiro trabalho que aborda a centralidade de tempo em GVTs. Os trabalhos anteriores estão mais interessados em qualificar a importância dos nós em uma rede do que em determinar os instantes de tempo mais importantes no contexto de um GVT.

A Centralidade de Tempo pode ser determinada de diferentes maneiras. Nesta obra, são apresentadas três métricas: *Grau do Tempo*, *Tempo de Cobertura* e *Centralidade de Cobertura*. A primeira é baseada na disposição das conexões da rede e as últimas são voltadas a processos de difusão de informação. As métricas desenvolvidas foram avaliadas em um conjunto de dados real, obtido de uma rede de contatos presenciais em um ambiente hospitalar. É importante ressaltar que as métricas de Tempo de Cobertura e Centralidade de Cobertura estão associadas a um problema bem específico, que é a difusão de informação/doença/comportamento como algo probabilístico.

Visto que muitos eventos modelados por GVTs são notoriamente cíclicos, a Centralidade de Tempo pode ser útil na construção de modelos de previsão de comportamento de GVTs, seja por padrões de estabelecimento de conexões, ou seja pelo aparecimento ou movimentação de nós ao longo do tempo. Com base em propriedades de instantes de tempo importantes no passado, uma vez que tais propriedades tornassem a ocorrer, o modelo de previsão em questão seria capaz de indicar instantes de tempo importantes em potencial.

Os resultados mostram que os instantes de tempo melhor avaliados, de acordo com as métricas desenvolvidas, podem realizar um processo de difusão mais rápido e eficiente. Quando os 5% melhores instantes de tempo são comparados a 5% de instantes de tempo escolhidos aleatoriamente, um processo de difusão pode atingir 20% da rede até 2,5 vezes mais rápido. Para atingir 60% da rede, o tempo de difusão da informação foi até 91% melhor. A cobertura da rede em 1500 passos do processo de difusão praticamente dobra com relação aos instantes de tempo escolhidos aleatoriamente. Para 2000 passos do processo de difusão, a cobertura da rede é cerca de 63% maior.

Também foi possível concluir que o tamanho do *snapshot*, no conjunto de dados avaliado, não influencia tão intensamente a classificação dos instantes de tempo. Há uma tendência dos instantes de tempo melhor qualificados implicarem na existência de instantes de tempo importantes quando o tamanho varia. A similaridade entre os classificações

dos 5% melhores instantes de tempo, segundo a métrica de Grau do Tempo, dos cenários analisados foi de cerca de 66%.

Como principais trabalhos futuros, pode-se definir e formalizar novas métricas para a avaliação da centralidade de instantes de tempo em GVTs. Como dito anteriormente, diferentes definições de centralidade de tempo podem surgir, levando em consideração diferentes noções de importância para instantes de tempo.

Além disso, é proposta a construção de modelos de previsão ou estimativa da Centralidade de Tempo em sistemas complexos que possam ser representados por GVTs. Dessa forma, é necessário desenvolver estratégias baseadas na monitoração e análise da dinâmica de GVTs para encontrar indícios no passado recente ou presente que indiquem que um determinado instante de tempo é o melhor (mais central) para, por exemplo, se iniciar uma difusão informação que seja mais eficiente sob algum aspecto. Intui-se que tais indícios podem ser encontrados nos padrões de estabelecimento de conexões ou no aparecimento ou movimentação de nós ao longo do tempo.

A identificação de instantes de tempo importantes demanda um conhecimento prévio do estado da rede e de sua evolução ao longo do tempo. Isto não é uma limitação, já que muitos eventos modelados por GVTs são notoriamente cíclicos, como a mobilidade em redes de transporte.

Outra possibilidade é a generalização do modelo de representação de GVTs para Grafos multi-aspectos. Os aspectos temporal (instantes de tempo) e estrutural (disposição dos nós e vértices), multicamadas e outras características independentes seriam considerados no processo de difusão de informação. Dessa forma, o conceito de centralidade pode ser estendido aos demais aspectos do grafo.

REFERÊNCIAS

- I. F. AKYILDIZ and M. C. VURAN. *Wireless sensor networks*, volume 4. John Wiley & Sons, 2010.
- A. BARRAT, M. BARTHELEMY, R. PASTOR-SATORRAS, and A. VESPIGNANI. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.*, 101:3747–3752, 2004.
- A. BARRAT, M. BARTHELEMY, and A. VESPIGNANI. *Dynamical Processes on Complex Networks*. Cambridge University Press, New York, NY, USA, 2008.
- D. BARTHOLOMEW. Stochastic models for social processes. In *J. Wiley*, New York, NY, USA, 1973.
- S. P. BORGATTI. Centrality and network flow. *Social Networks*, 27(1):55–71, 2005.
- B. BUI-XUAN, A. FERREIRA, and A. JARRY. Evolving graphs and least cost journeys in dynamic networks. In *Proceedings of WiOpt'03 – Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks*, pages 141–150, Sophia Antipolis, March 2003. INRIA Press.
- A. CASTEIGTS, P. FLOCCHINI, W. QUATTROCIOCCHI, and N. SANTORO. Time-varying graphs and dynamic networks. In *Proceedings of the 10th International Conference on Ad-hoc, Mobile, and Wireless Networks, ADHOC-NOW'11*, pages 346–359, Berlin, Heidelberg, 2011. Springer-Verlag.
- J. H. CHOI, G. BARNETT, and B. CHON. Comparing world city networks: a network analysis of internet backbone and air transport intercity linkages. *Global Networks*, 6(1):81–99, 2006.
- D. EASLEY and J. KLEINBERG. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.
- J. R. EVANS and E. MINIEKA. *Optimization Algorithms for Networks and Graphs, Second Edition*,. CRC, 2 edition, 1992.
- A. FERREIRA. Building a reference combinatorial model for manets. *IEEE Network*, 18(5):24–29, 2004.
- L. C. FREEMAN. A Set of Measures of Centrality Based on Betweenness. *Sociometry*, 40(1):35–41, March 1977.
- L. C. FREEMAN. Centrality in social networks conceptual clarification. *Social Networks*, page 215, 1978.

- S. GHOSH. *Distributed Systems: An Algorithmic Approach*. Chapman & Hall/CRC computer and information science series. Chapman & Hall/CRC, Boca Raton, London, New York, 2007.
- P. HOLME. Network reachability of real-world contact sequences. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 71(4), 2005.
- P. HOLME and J. SARMAKI. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.
- T. IDEKER, V. THORSSON, J. A. RANISH, R. CHRISTMAS, J. BUHLER, J. K. ENG, R. BUMGARNER, D. R. GOODLETT, R. AEBERSOLD, and L. HOOD. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–934, 2001.
- L. ISELLA, M. ROMANO, A. BARRAT, C. CATTUTO, V. COLIZZA, W. VAN DEN BROECK, F. GESUALDO, E. PANDOLFI, L. RAV, C. RIZZO, and A. E. TOZZI. Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLOS ONE*, 6(2):e17144, 02 2011.
- T. ITO, T. CHIBA, R. OZAWA, M. YOSHIDA, M. HATTORI, and Y. SAKAKI. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98(8):4569–4574, 2001.
- A. KHELIL, C. BECKER, J. TIAN, and K. ROTHERMEL. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '02, pages 54–60, New York, NY, USA, 2002. ACM.
- H. KIM and R. ANDERSON. Temporal node centrality in complex networks. *Phys. Rev. E*, 85:026107, Feb 2012.
- V. KOSTAKOS. Temporal graphs. *Physica A: Statistical Mechanics and its Applications*, (6):1007–1023, March 2009.
- F. LEWIS. Wireless sensor networks. *Smart environments: technologies, protocols, and applications*, pages 11–46, 2004.
- S. LI, C. M. ARMSTRONG, N. BERTIN, H. GE, S. MILSTEIN, M. BOXEM, P. VIDALAIN, J. J. HAN, A. CHESNEAU, T. HAO, D. S. GOLDBERG, N. LI, M. MARTINEZ, J. F. RUAL, P. LAMESCH, L. XU, M. TEWARI, S. L. WONG, L. V. ZHANG, G. F. BERRIZ, L. JACOTOT, P. VAGLIO, J. REBOUL, T. HIROZANE-KISHIKAWA, Q. LI, H. W. GABEL, A. ELEWA, B. BAUMGARTNER, D. J. ROSE, H. YU, S. BOSAK, R. SEQUERRA, A. FRASER, S. E. MANGO, W. M. SAXTON, S. STROME, S. VAN

- DEN HEUVEL, F. PIANO, J. VAN DEN HAUTE, C. SARDET, M. GERSTEIN, L. DOUCETTE-STAMM, K. C. GUNSALUS, J. W. HARPER, M. E. CUSICK, F. P. ROTH, D. E. HILL, and M. VIDAL. A map of the interactome network of the metazoan *c. elegans*. *Science*, 303(5657):540–543, 2004.
- L. LOVASZ. Random walks on graphs: A survey. In *Combinatorics, Paul Erdos is Eighty*, volume 2, pages 353–398. Janos Bolyai Mathematical Society, Budapest, 1996.
- J. C. LUCET, C. LAOUENAN, G. CHELIUS, N. VEZIRIS, D. LEPELLETIER, A. FRIGGERI, D. ABITEBOUL, E. BOUVET, F. MENTRE, and E. FLEURY. Electronic Sensors for Assessing Interactions between Healthcare Workers and Patients under Airborne Precautions. *PLoS ONE*, 7(5):e37893, May 2012.
- M. E. J. NEWMAN. A measure of betweenness centrality based on random walks. *Social networks*, 27(1):39–54, 2005.
- M. E. J. NEWMAN. Mathematics of networks. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.
- V. NICOSIA, J. TANG, M. MUSOLESI, G. RUSSO, C. MASCOLO, and V. LATORA. Components in time-varying graphs. *CoRR*, abs/1106.2134, 2011.
- V. NICOSIA, J. TANG, C. MASCOLO, M. MUSOLESI, G. RUSSO, and V. LATORA. Graph metrics for temporal networks. *CoRR*, abs/1306.0493, 2013.
- T. OPSAHL, AGNEESSENS F., and J. SKVORETZ. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.
- R. K. PAN and J. SARMAKI. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E*, 84:016105, Jul 2011.
- G. SABIDUSSI. The centrality index of a graph. *Psychometrika*, 31(4):581–603, December 1966.
- N. SANTORO, W. QUATTROCIOCCHI, P. FLOCCHINI, A. CASTEIGTS, and F. AMBLARD. Time-varying graphs and social network analysis: Temporal indicators and metrics. *CoRR*, abs/1102.0629, 2011.
- B. SHULGIN, L. STONE, and Z. AGUR. Pulse vaccination strategy in the sir epidemic model. *Bulletin of Mathematical Biology*, 60:1123–1148, 1998.
- M. STARNINI, A. BARONCHELLI, A. BARRAT, and R. PASTOR-SATORRAS. Random walks on temporal networks. *Phys. Rev. E*, 85, May 2012.

- J. STEHL, N. VOIRIN, A. BARRAT, C. CATTUTO, V. COLIZZA, L. ISELLA, C. REGIS, J.-F. PINETON, N. KHANAFER, W. VAN DEN BROECK, and P. VANHEMS. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(87), July 2011a.
- J. STEHL, N. VOIRIN, A. BARRAT, C. CATTUTO, L. ISELLA, J.-F. PINETON, M. QUAGGIOTTO, W. VAN DEN BROECK, C. REGIS, B. LINA, and P. VANHEMS. High-resolution measurements of face-to-face contact patterns in a primary school. *CoRR*, abs/1109.1015, 2011b.
- J. TANG, M. MUSOLESI, C. MASCOLO, and V. LATORA. Temporal distance metrics for social network analysis. In *Proceedings of the 2Nd ACM Workshop on Online Social Networks*, WOSN '09, pages 31–36, New York, NY, USA, 2009. ACM.
- J. TANG, M. MUSOLESI, C. MASCOLO, V. LATORA, and V. NICOSIA. Analysing information flows and key mediators through temporal centrality metrics. In *Proceedings of the 3rd Workshop on Social Network Systems*, SNS '10, pages 3:1–3:6, New York, NY, USA, 2010a. ACM. ISBN 978-1-4503-0080-3.
- J. TANG, S. SCELLATO, M. MUSOLESI, C. MASCOLO, and V. LATORA. Small-world behavior in time-varying graphs. *Physical Review E*, 81(5), May 2010b.
- J. WANG, H. MO, F. WANG, and F. JIN. Exploring the network structure and nodal centrality of china's air transport network: A complex network approach. *Journal of Transport Geography*, 19(4):712–721, 2011.
- K. WEHMUTH, A. ZIVIANI, and E. FLEURY. A unifying model for representing time-varying graphs. *CoRR*, abs/1402.3488, 2014a.
- K. WEHMUTH, A. ZIVIANI, and E. FLEURY. On multiaspect graphs. *CoRR*, abs/1408.0943, 2014b.
- T. YASSERI, R. SUMI, and J. KERTESZ. Circadian patterns of wikipedia editorial activity: A demographic analysis. *CoRR*, abs/1109.1746, 2011.