

Centralidade de Tempo em Grafos Variantes no Tempo

Eduardo Chinelate Costa¹, Alex Borges Vieira¹, Ana Paula Couto da Silva²

¹ Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Juiz de Fora – Juiz de Fora – MG

²Departamento de Ciência da Computação
Universidade Federal de Minas Gerais – Belo Horizonte – MG

Abstract. *Centrality usually refers to metrics that assess the relative importance of vertices. However, in Time-Varying Graphs (TVGs) it is possible to assess the importance of time instants (or states) of a graph throughout its existence. Determining important time instants may be useful to defining best times to spread, generate models and predict the behavior of TVGs. In this paper, we define time centrality in TVGs. Time centrality evaluates the relative importance of time instants. We present and evaluate two time centrality metrics focused on information dissemination processes. Our results show that the best classified time instants, according to created metrics, can make the diffusion process up to 2.5 times faster and achieve twice the number of nodes in certain cases.*

Resumo. *O conceito de centralidade geralmente se refere a métricas que avaliam a importância relativa dos vértices. Entretanto, Grafos Variantes no Tempo (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 pode ser útil para definir melhores momentos para difusão, gerar modelos e prever o comportamento de GVTs. Neste trabalho, é definido o conceito de Centralidade de Tempo em GVTs. A centralidade de tempo avalia a importância relativa dos instantes de tempo. São apresentadas e avaliadas duas métricas de centralidade de tempo voltadas a processos de difusão de informação. Os resultados mostram que os instantes de tempo melhor classificados, de acordo com as métricas criadas, podem tornar o processo de difusão até 2,5 vezes mais rápido e atingir o dobro do número de nós em certos casos.*

1. Introdução

Grafos variantes no tempo (GVTs) ganharam interesse recente, em especial por contemplar a evolução temporal, tanto de nós, quanto de arestas. Nesse novo cenário, de maneira similar a grafos estáticos, o conceito de centralidade geralmente se refere a métricas que avaliam a importância relativa dos vértices. Não obstante, pesquisadores sempre tentaram mapear o conceito do cenário tradicional para o novo.

Entretanto, GVTs possibilitam uma nova abordagem para se determinar centralidade. De fato, como agora é possível acompanhar a evolução do grafo, não só vértices e arestas são importantes, mas o tempo se torna um novo personagem a ser avaliado. Assim, uma nova maneira de se determinar centralidade em grafos é proposta. Neste trabalho, a *Centralidade de Tempo em Grafos Variantes no Tempo* é definida e formalizada, avaliando a importância dos instantes de tempo (ou estados) de um grafo ao longo de sua existência.

Determinar instantes de tempo importantes nesse contexto pode ter fortes aplicações práticas. Em particular, pode ser útil para definir melhores momentos para difusão de uma informação, gerar modelos e prever o comportamento de GVTs.

Além da formalização de Centralidade de Tempo, são apresentadas duas métricas voltadas a processos de difusão de informação em uma rede. As métricas foram avaliadas em um conjunto de dados real e os resultados indicam que, se os instantes de tempo com maior centralidade fossem usados para difundir algo na rede, tal processo seria até 2,5 vezes mais rápido. Caso houvesse limitações de tempo, os resultados também indicam um melhor desempenho da difusão com a escolha do instante de tempo apropriado. Nesse caso, a cobertura da difusão é até 2 vezes maior, quando comparado a uma escolha aleatória do momento de início da difusão.

2. Objetivos e Contribuições

Até onde se sabe, 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 algumas aplicações modeladas em GVTs, 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. Decidir *quando*, e não somente *onde* iniciar um processo de difusão pode ser de extrema importância para um resultado mais eficaz.

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. Por sua vez, [Spasojevic et al. 2015] mostrou recentemente evidências de que recomendar melhores tempos para um usuário postar em redes sociais aumenta a probabilidade de obtenção de respostas do público.

Visto que muitos eventos modelados por GVTs são notoriamente cíclicos, como mobilidade em redes de transportes e surtos epidêmicos de doenças, a Centralidade de Tempo seria útil na construção de modelos de previsão. As estratégias para a possível identificação antecipada de instantes de tempo centrais podem basear-se na descoberta de evidências no passado recente ou no momento presente da evolução do GVT, indicando que um determinado instante de tempo é relativamente melhor (isto é, mais central) com relação a um conjunto de outros.

Os resultados experimentais mostram que processos de difusão de informação iniciados em instantes de tempo melhor avaliados pelas métricas de centralidade de tempo — mais centrais — são efetivamente mais rápidos e/ou mais abrangentes que processos iniciados nos demais instantes de tempo. 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 um dado número de passos do processo de difusão praticamente dobra com relação aos instantes de tempo escolhidos aleatoriamente.

3. Centralidade de Tempo

Assim como na Centralidade de Nó, diferentes definições de Centralidade de Tempo podem refletir diferentes noções de importância para instantes de tempo visando aplicações

distintas. Em outras palavras, da mesma forma que a Centralidade de Nó cria um ranqueamento dos nós sob um critério de importância arbitrária, a Centralidade de Tempo cria um ranking de instantes de tempo usando algum critério para medir a importância relativa que um instante de tempo pode ter quando comparado com os outros.

São apresentadas duas métricas de centralidade de tempo que podem avaliar a difusão em GVTs a partir de diferentes perspectivas. Para definir tais métricas, um GVT genérico $H = (V, E, T)$ [Wehmuth et al. 2014] é considerado. Seja $N = |T(H)|$ o tamanho do conjunto $T(H)$. Cada instante de tempo individual $t_i \in T(H)$, onde $0 \leq i < N$, de modo que $T(H) = \{t_0, t_1, \dots, t_{N-1}\}$.

3.1. Tempo de Cobertura

A métrica de *Tempo de Cobertura (Cover Time (CT))* do ponto de vista da centralidade de tempo avalia o número mínimo de instantes de tempo que uma difusão leva para atingir uma dada fração dos nós. Mais formalmente, o tempo de cobertura $CT(t_i, \tau)$ retorna a quantidade média de instantes de tempo entre todos os nós para que uma difusão iniciada no tempo t_i alcance uma certa fração τ da rede, ou seja,

$$CT(t_i, \tau) = \begin{cases} \frac{1}{|V(H)|} \sum_{u \in V(H)} d_t(t_i, u, \tau), & \text{Se } \tau \text{ é atingido,} \\ \infty, & \text{Caso contrário.} \end{cases} \quad (1)$$

onde $d_t(t_i, u, \tau)$ é o número de instantes de tempo para que uma difusão iniciada no nó u no instante de tempo t_i alcance a fração τ de nós. Se a fração τ de nós não é alcançada pelo processo de difusão iniciado em qualquer nó em t_i , então $CT(t_i, \tau)$ é considerado ∞ .

O processo de difusão segue uma busca em largura (BFS), ou seja, um nó genérico (u, t_i) inicia a difusão por meio de envio de informações para todos os seus nós adjacentes. Em seguida, os nós temporais adjacentes retransmitem a informação para os seus próprios nós adjacentes no próximo instante de tempo, e assim por diante. A informação é distribuída até que a fração τ de nós seja atingida e o tempo necessário para a difusão é armazenado. Esta avaliação é repetida para cada nó $(u, t_i) \in VT(H)$. O tempo médio de difusão é a centralidade de tempo de cobertura do instante de tempo t_i .

3.2. Centralidade de Cobertura

A métrica de *Centralidade de Cobertura (Time-Constrained Coverage (TCC))* do ponto de vista da Centralidade de Tempo avalia a cobertura do GVT atingida por um processo de difusão depois de um número limitado de passos. Aqui é assumido o mesmo processo de difusão definido para o tempo de cobertura (Seção 3.1). Mais precisamente, para um dado instante de tempo t_i , $TCC(t_i, \Phi)$ retorna a fração média de nós do TVG atingidos pelo processo de difusão em Φ passos. Mais formalmente,

$$TCC(t_i, \Phi) = \frac{1}{|V(H)|^2} \sum_{u \in V(H)} d_c(t_i, u, \Phi), \quad (2)$$

onde $d_c(t_i, u, \Phi)$ é o número de nós atingidos a partir do nó u depois de Φ passos quando a difusão começa no tempo t_i . Novamente, é importante lembrar que a métrica de centralidade de cobertura considera uma média dos processos de difusão iniciados a partir de cada nó.

4. Avaliação da Centralidade de Tempo

As métricas de centralidade de tempo propostas foram avaliadas usando um conjunto de dados coletados no âmbito do projeto MOSAR [Lucet et al. 2012]. O MOSAR (*Mastering hOSpital Antimicrobial Resistance and its spread*) é um projeto de colaboração científica que compreende várias instituições de pesquisa, que se concentram na dinâmica de transmissão de bactérias resistentes aos antimicrobianos (AMRB) em ambientes de alto risco (unidades de terapia intensiva e centros cirúrgicos). O conjunto de dados adotado consiste no registro de contatos presenciais (de médicos, enfermeiros, funcionários e pacientes) em uma determinada ala hospitalar por um período de duas semanas. Cada um dos 160 voluntários que participaram no estudo foi equipado com um dispositivo de RFID que detectou a presença de outros dispositivos dentro de uma distância pequena (cerca de um metro). Os dispositivos de identificação eram únicos e sempre associados à mesma pessoa. A cada 30 segundos durante o período de duas semanas, cada dispositivo registrou a lista de todos os dispositivos (nós) que estavam dentro de sua área de cobertura, a fim de estabelecer o arranjo dos contatos entre eles (arestas).

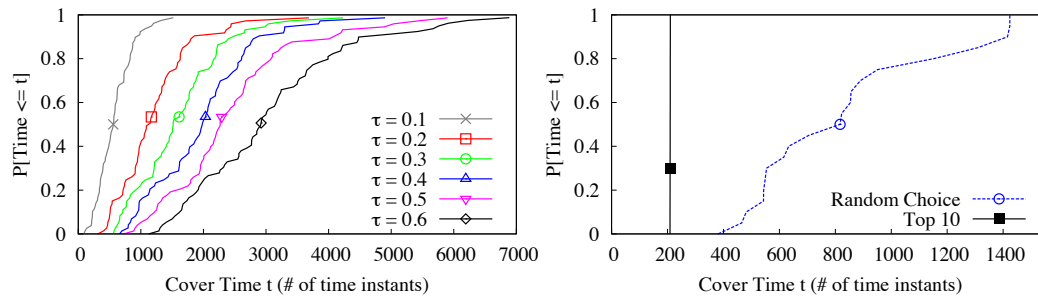
O conjunto de dados MOSAR é modelado como um GVT $M = (V, E, T)$, conforme descrito em [Wehmuth et al. 2014], onde $V(M)$ é o conjunto de todos os participantes, $E(M)$ é o conjunto de todos os contatos interpessoais, e $T(M)$ é o conjunto de instantes de tempo sequenciais com 30 segundos de granularidade. Dessa forma, $|V(M)| = 160$ e $|T(M)| = 40320$. Os contatos são representados por arestas não-orientadas, ou seja, a aresta entre duas pessoas é representada por uma aresta orientada e sua recíproca.

4.1. Tempo de Cobertura

Inicialmente são mostrados os resultados para a métrica de Tempo de Cobertura (Figura 1). Vale ressaltar que menores valores de tempo de cobertura indicam instantes de tempo mais centrais, que podem difundir uma informação em tempo.

A Figura 1(a) apresenta a função de distribuição acumulada (CDF) do tempo de cobertura para diferentes valores da fração τ de nós atingidos no GVT MOSAR. Seis diferentes cenários são considerados com $\tau = \{0.1, 0.2, \dots, 0.6\}$. Os resultados apresentam os primeiros 33000 instantes de tempo, para que haja espaço para a difusão se espalhar ao longo dos 40320 instantes de tempo do GVT MOSAR.

Cada valor em uma curva representa a fração dos instantes de tempo do GVT que alcançam a correspondente fração τ de nós em, no máximo, t instantes de tempo. A fim melhorar a visualização dos dados, os símbolos estão em posições arbitrárias nas curvas (o mesmo acontece com as demais figuras). Por exemplo, para atingir 60% do GVT MOSAR ($\tau = 0.6$), 20% dos instantes de tempo leva menos de 2000 instantes de tempo. Em outras palavras, a difusão da informação iniciada em até 20% dos instantes de tempo leva menos de 16 horas e 40 minutos para atingir 60% dos nós no GVT MOSAR, somente difundindo de maneira oportunista através do contato presencial entre os indivíduos.



(a) Distribuição do Tempo de Cobertura

(b) 10 Melhores x Escolha Aleatória

Figura 1. Função de Distribuição Acumulada do Tempo de Cobertura na rede MOSAR

Na Figura 1(b), para o GVT MOSAR, é comparado o desempenho dos 10 melhores instantes de tempo de acordo com a métrica de tempo de cobertura contra 10 instantes de tempo escolhidos aleatoriamente para iniciar uma difusão. Tomando o resultado médio como referência, para atingir 10% dos nós ($\tau = 0.1$) no GVT MOSAR, o tempo de cobertura foi de 266 instantes de tempo quando a difusão começa nos 10 melhores instantes de tempo; enquanto que o tempo de cobertura foi de 680 instantes de tempo, quando a difusão começa nos 10 instantes de tempo escolhidos aleatoriamente. Ou seja, considerando os instantes de tempo mais centrais, a difusão leva 2.5 vezes menos tempo quando comparado com a difusão iniciada em instantes de tempo aleatórios.

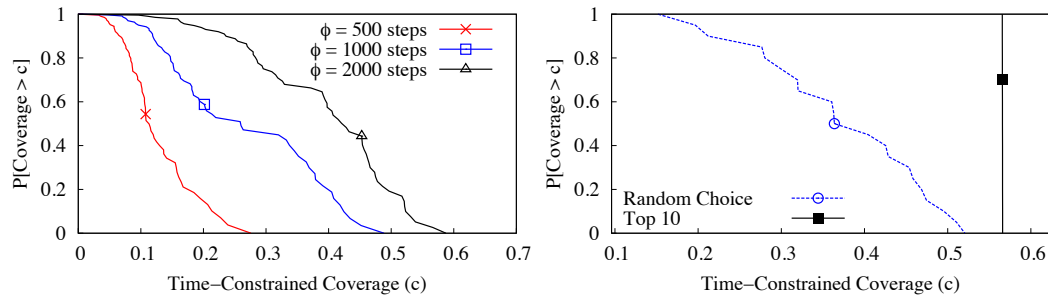
4.2. Centralidade de Cobertura

A centralidade de cobertura também é capaz de selecionar instantes de tempo que apresentem desempenhos de difusão distintos. A Figura 2(a) apresenta a função de distribuição acumulada complementar da centralidade de cobertura do GVT MOSAR para diferentes valores de Φ passos. Assim como no tempo de cobertura, os primeiros 33000 instantes de tempo são analisados.

Cada valor em uma curva representa a fração dos instantes de tempo do GVT que alcançam a correspondente centralidade de cobertura c em, no mínimo, Φ passos. Por exemplo, com um limite de $\Phi = 2000$ passos, a difusão iniciada em 80% dos instantes de tempo alcança no mínimo 28% dos nós no GVT MOSAR, observando que a máxima cobertura alcançada é de 58%.

Conforme esperado, quanto maior o número de passos Φ , maior a cobertura global da rede. Vale ressaltar, contudo, que em todos os casos, um número muito limitado de instantes de tempo — na verdade, os mais centrais — foram eficientemente superiores na cobertura do GVT MOSAR, com um número limitado de passos, quando a difusão foi iniciada neles. Na Figura 2(b), são comparados o desempenho dos 10 melhores instantes de tempo — os mais centrais — de acordo com a métrica de centralidade de cobertura, e o desempenho dos 10 instantes de tempo escolhidos aleatoriamente para iniciar uma difusão de informação. De fato, iniciar uma difusão em qualquer um dos 10 melhores instantes de tempo alcançaria, no mínimo, 56% da cobertura da rede, enquanto que nos

instantes de tempo escolhidos aleatoriamente, o processo de difusão alcançaria não mais que 33% dos nós do GVT MOSAR, considerando um limite de $\Phi = 2000$ passos.



(a) Distribuição da Centralidade de Cobertura

(b) 10 Melhores x Escolha Aleatória

Figura 2. Função de Distribuição Acumulada da Centralidade de Cobertura na rede MOSAR

5. Conclusões

Na dissertação, foi introduzido o conceito de Centralidade de Tempo em GVTs. A centralidade de tempo avalia a importância relativa de um instante de tempo dentro do contexto de um GVT. Além disso, foram apresentadas duas métricas de centralidade de tempo, baseadas em processos de difusão, e ambas foram avaliadas usando um conjunto de dados real, que representa uma rede dinâmica de contato interpessoal. Os resultados mostram que iniciar uma difusão em um instante de tempo mais central, de acordo com nossas métricas, pode resultar em um processo de difusão mais rápido e abrangente.

Como trabalho futuro, é cogitado o desenvolvimento de modelos de previsão baseados na centralidade de tempo para sistemas complexos que podem ser representados por GVTs. Para isso, é necessário elaborar estratégias que possibilitem a identificação de instantes de tempo mais centrais, baseadas em evidências no passado recente ou no presente da evolução do GVT. Dessa forma, é possível indicar prováveis instantes de tempo centrais à medida em que eles surgirem.

Referências

- Lucet, J. C., Laouenan, C., Chelius, G., Veziris, N., Lepelletier, D., Friggeri, A., Abiteboul, D., Bouvet, E., Mentre, F., and Fleury, E. (2012). Electronic Sensors for Assessing Interactions between Healthcare Workers and Patients under Airborne Precautions. *PLoS ONE*, 7(5):e37893.
- Spasojevic, N., Li, Z., Rao, A., and Bhattacharyya, P. (2015). When-to-post on social networks. In *Prof. of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM KDD)*.
- Wehmuth, K., Ziviani, A., and Fleury, E. (2014). A Unifying Model for Representing Time-Varying Graphs. *ArXiv e-prints*.