

Caracterização de Mobilidade e Detecção de Comunidades baseadas em Tópicos de Interesse

Iran F. Ribeiro¹, Lucas Castanheira², Alberto Schaeffer-Filho²,
Weverton Cordeiro², Vinícius F. S. Mota^{1*}

¹ Departamento de Informática – Universidade Federal do Espírito Santo
Vitória – Brasil

iran.ribeiro@aluno.ufes.br, vinicius.mota@inf.ufes.br

²Instituto de Informática – Universidade Federal do Rio Grande do Sul
Porto Alegre – Brasil

{lbcastanheira, alberto, weverton.cordeiro}@inf.ufrgs.br

Abstract. *Mobile ad hoc networks pose challenges such as connection setup, routing, and network lifetime due to node mobility. As such, researchers have been analyzing mobility aspects, e.g., contact and intercontact time between devices and the communities based on these contacts, in scenarios such as university campuses, vehicles, and conferences. In this paper, we characterize the mobility of conference participants and propose community detection based on topics of interest. The data used in this study, collected during SBRC 2019 using a gamification app, was anonymized and modeled as a temporal graph. Results show that, although each participant has a high number of contacts, contact time is low. Besides, communities based on topics of interest are similar to k -cliques of the graphs.*

Resumo. *Redes móveis ad hoc impõem desafios como estabelecimento de conexões, descoberta de rotas e duração da rede devido à mobilidade dos nós. Nesse sentido, pesquisadores vem analisando aspectos de mobilidade, p.ex., tempo de contato e entre-contato dos dispositivos e as comunidades formadas, em cenários como campus universitário, veículos e em conferências. Este trabalho caracteriza a mobilidade de participantes de uma conferência e propõe a detecção de comunidades baseadas em tópicos de interesse. Os dados usados nesse estudo, coletados durante o SBRC 2019 por meio de um app de gamificação, foram anonimizados e modelados como um grafo temporal. Os resultados mostram que, embora cada participante tenha um alto número de contatos, o tempo de contato é baixo. Além disso, as comunidades detectadas a partir dos tópicos de interesse são similares aos k -cliques do grafo temporal.*

1. Introdução

A caracterização de encontros entre dispositivos móveis é de suma importância para diversos tipos de redes móveis, tais como redes de sensores móveis, redes veiculares e comunicação dispositivo-a-dispositivo (D2D) [Mota et al. 2014]. A importância

*Os autores agradecem à CAPES, CNPq e FAPES pelo apoio financeiro parcial a este trabalho.

deve-se ao fato desses tipos de redes, também chamadas de redes oportunísticas, serem caracterizadas pela transferência de dados entre dispositivos móveis quando ocorre uma oportunidade de contato entre estes. No contexto da próxima geração de redes móveis, chamada 5G, a comunicação D2D tem o papel de descarregar o tráfego da rede infraestruturada, estabelecendo conexões e movendo o tráfego de dado para dispositivos próximos entre si [Höyhty et al. 2018]. No entanto, a qualidade de serviço da comunicação D2D é influenciada diretamente pelo comportamento humano e o padrão de encontro entre os dispositivos móveis que compõem a rede.

Na avaliação de redes móveis, modelos de mobilidade sintéticos são comumente utilizados, o que pode levar a má interpretação dos resultados e ao questionamento sobre o quão realístico é o modelo utilizado. Para contornar este problema, diversos estudos utilizaram *traces* contendo dados de rede de telefonia [Marques-Neto et al. 2018], dados de redes sociais [Akabane et al. 2019] ou mesmo realizaram experimentos utilizando dispositivos específicos em eventos e em conferências [Pietiläinen and Diot 2012]. O estudo de mobilidade e contatos reais pode esclarecer onde, quando e quanto tempo durará o encontro entre dois ou mais dispositivos, auxiliando no desenvolvimento de protocolos de disseminação de dados mais eficientes.

Nesse trabalho, apresentamos uma análise de comunidades temporais com base em interesses em eventos científicos, utilizando como estudo de caso dados coletados durante a 37ª edição do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2019). Os dados foram obtidos através de um sistema de gamificação desenvolvido para incentivar a participação e o engajamento nas sessões técnicas e nos eventos satélites. Assumindo que os dispositivos móveis dos congressistas em uma mesma sala poderiam conectar-se, um resultado secundário do sistema desenvolvido – e principal alvo deste estudo – é a caracterização dos contatos entre os participantes nas sessões técnicas. Para isso, os dados foram anonimizados e a dinâmica da rede formada pelos participantes nas sessões foi modelada como um grafo de contato temporal, no qual os vértices representam os participantes e as arestas representam os contatos.

As principais características de mobilidade avaliadas são o número de contatos, a duração dos contatos (entre participantes e instituições) e o tempo entre reencontros dos participantes. Adicionalmente, a detecção de comunidades é importante para compreensão da estrutura social no cenário analisado [Wang et al. 2016]. Diferentemente dos trabalhos tradicionais que utilizam as características dos contatos estabelecidos na rede para detecção de comunidades, este trabalho considera os tópicos de interesse para inferir as comunidades. Desta forma, as principais contribuições deste trabalho são: (i) a caracterização da mobilidade dos participantes entre sessões técnicas de um evento científico; (ii) a detecção de comunidades com base em tópicos de interesse; e (iii) a caracterização da mobilidade entre as sessões das comunidades detectadas.

O restante do artigo está organizado como segue. A Seção 2 revisa os trabalhos relacionados, enquanto que a Seção 3 discute o estudo de caso. A Seção 4 apresenta o modelo de rede utilizado, seguido pela caracterização dos contatos entre os participantes na Seção 5. Os impactos dos resultados em redes oportunísticas são discutidos na Seção 6. A Seção 7 fecha o artigo com as considerações finais.

2. Trabalhos Relacionados

Redes sociais móveis ou oportunísticas têm sido amplamente estudadas na literatura [Wang et al. 2016, Zhao and Song 2017, Awuor et al. 2018, Wu et al. 2019]. As motivações vão desde o potencial de adoção das mesmas para computação [Trifunovic et al. 2017] e *caching* [Wu et al. 2019] oportunísticos, bem como permitir reuso de recursos de redes celulares visando maior eficiência no uso do espectro [Wang et al. 2016]. Um consenso na literatura é de que o sucesso desse tipo de aplicação depende do entendimento sobre como redes sociais oportunísticas se organizam e quais suas principais propriedades [Zhao and Song 2017]. Em paralelo, diversos autores têm investigado as potencialidades de redes oportunísticas na disseminação epidêmica de conteúdos em redes 5G [Dede et al. 2017], DTN [McGeehan et al. 2019], redes centradas em conteúdos [Appel et al. 2019], etc.

Recentes avanços teóricos e práticos nessa área têm estimulado pesquisadores a investigarem o conjunto de dados disponíveis de interações humanas e a revisitarem teorias sobre como essas redes se formam, suas principais propriedades, e como elas podem ser exploradas para disseminação de conteúdos [Viard et al. 2015]. Pietiläinen e Diot [Pietiläinen and Diot 2012], por exemplo, usaram como base em seu estudo ambientes acadêmicos (encontros científicos e campi universitários) e concluíram que, embora a eficiência na disseminação de conteúdos ocorra principalmente devido a nodos com grandes taxas de contato entre si, o processo de disseminação é prejudicado se esses nodos tornam-se fortemente engajados em comunidades temporárias. Mais recentemente, Viard *et al.* [Viard et al. 2015] identificaram novos padrões de contatos entre pessoas em diferentes escalas temporais, em um experimento em uma escola de ensino médio. Em comum a essas pesquisas, observa-se não apenas o interesse em entender como as interações sociais ocorrem (e mesmo como elas são influenciadas pela tecnologia à disposição [Trifunovic et al. 2017]), mas também como a tecnologia pode apoiar essas interações [Awuor et al. 2018].

Uma lacuna na literatura refere-se à análise de redes sociais oportunísticas influenciada por tópicos de interesse. Nesse contexto, o presente trabalho visa analisar os dados obtidos de interações em uma conferência acadêmica, e analisar padrões de comportamentos e detecção de comunidades formadas pelos participantes dessa conferência. Em especial, procura-se verificar aspectos como a distribuição do grau de nodos, do tempo de contato e entre contatos e as comunidades formadas pela interseção de sessões técnicas visitadas, isto é, participantes com os mesmos interesses, comparando-as com algoritmos tradicionais de detecção de comunidades.

3. Estudo de Caso: SBRC 2019

Apresentaremos a seguir um breve resumo do evento científico onde os dados foram coletados assim como uma descrição do sistema desenvolvido para obter esses dados.

3.1. Descrição do Evento e Sistema de Gamificação

A análise apresentada nesse trabalho é baseada em dados obtidos durante a 37ª edição do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2019), que ocorreu entre os dias 6 e 10 de maio de 2019, em Gramado-RS¹. O SBRC 2019 contou com 782 inscritos de acordo com o sistema ECOS da SBC.

¹As sessões técnicas e workshops podem ser consultados em <http://sbrc2019.sbc.org.br/>

Para o SBRC 2019, um sistema de gamificação foi concebido com o objetivo principal de incentivar os usuários a participarem das sessões técnicas do evento, e oferecer incentivos a perguntas ao final das apresentações. O objetivo secundário para o desenvolvimento de tal sistema foi o de permitir a coleta de dados para posterior análises e correlações de interesses e participações da comunidade.

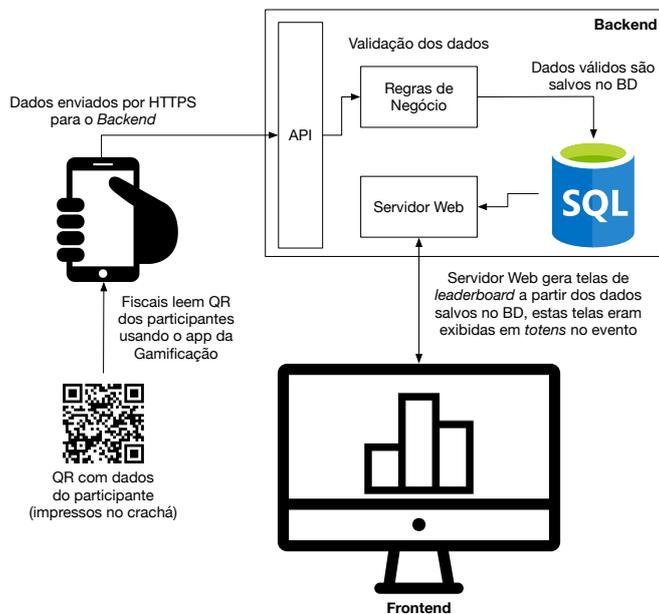


Figura 1. Arquitetura do Sistema de Gamificação do SBRC 2019

Os principais módulos do sistema de gamificação são descritos a seguir. Participantes do evento receberam crachás com QR code únicos, que eram lidos pela equipe da organização local através de aplicativos iOS/Android. A leitura dos QR codes permitia a contabilização de entradas nas sessões técnicas e de perguntas realizadas pelos participantes. As saídas, entretanto, não eram registradas. Um servidor consolidava a pontuação dos participantes, e *leaderboards* espalhados pelo centro de eventos mostravam os maiores pontuadores. Uma visão geral da arquitetura do sistema de gamificação é ilustrada na Figura 1.

App iOS/Android: Utilizado para leitura dos QR codes e envio ao servidor. O app foi distribuído entre os monitores das sessões técnicas, sendo utilizado para registrar o *check-in* dos participantes e as perguntas feitas nas sessões técnicas.

Backend Python: Utilizado para receber os dados dos apps e validar as regras de negócio da gamificação do evento (*e.g.*, pontos por participação em sessão técnica, evitar que usuário pontue mais de uma vez em uma sessão, etc). Os dados eram armazenados em um banco de dados MySQL, após normalização e validação.

Frontend Python: Utilizado na geração dos *leaderboards*, exibindo os *top-5* participantes, instituições, e sessões técnicas. Resultados parciais eram exibidos em três *totens* digitais disponibilizados no centro de eventos.

Destaca-se que a participação na gamificação era opcional. Os participantes poderiam não realizar a leitura de seu QR code, ou solicitar *opt-out* do sistema com um dos membros da equipe local. Assim, a análise apresentada nesse artigo corres-

ponde somente aos dados coletados. Ressalta-se que os dados foram anonimizados.

3.2. Sumarização e Filtragem dos Dados Coletados

Foram coletados e armazenados pelo sistema de gamificação 2.551 registros, contendo as entradas nas Sessões Técnicas (STs), *Workshops*, *Keynotes* e Minicursos, além das perguntas feitas pelos participantes. Neste trabalho, nas análises realizadas a partir da Seção 5, foram utilizados apenas os dados referentes aos registros de entradas nas diversas sessões técnicas e nos eventos satélites.

Tabela 1. Dados coletados pelo aplicativo de gamificação durante o SBRC 2019

Registro de entrada nas STs	1.481
Registros removidos	1070
Total participantes	458
Participantes nas Sessões	414
Instituições nas Sessões	102

A Tabela 1 sumariza os dados coletados. No total, participaram da gamificação 458 dos 782 inscritos, de 102 instituições distintas. Removemos 1070 registros que apresentavam dados incoerentes, faltantes ou que não faziam parte das sessões técnicas e eventos satélites. Por exemplo, registros em que os participantes foram inseridos em uma data diferente do dia da ST e situações em que participantes eram inseridos em uma ST após a mesma ter terminado. Em alguns casos, os campos referentes ao estado de origem e instituição ao qual alguns participantes pertenciam estavam trocados e foram corrigidos manualmente.

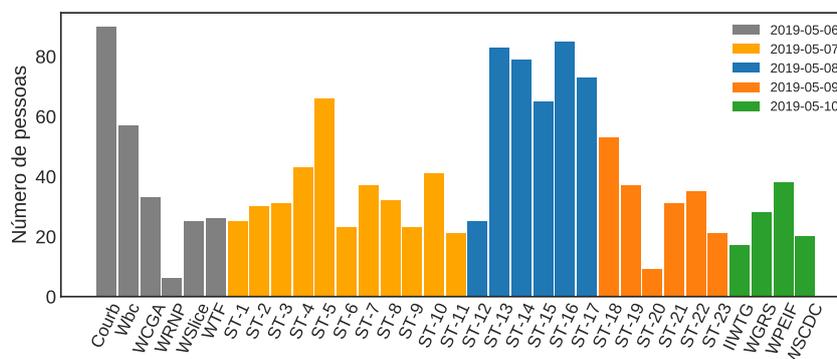


Figura 2. Número de pessoas em cada sessão

A Figura 2 apresenta o número de participantes por sessões técnicas e workshops. Ocorreram 6 e 4 workshops no primeiro e último dia do evento, respectivamente. As sessões de ST-1 à ST-11 ocorreram durante o segundo dia, as próximas 6 sessões (ST-12 à ST-17) ocorreram no terceiro dia, e as 6 últimas sessões (ST-18 à ST-23) aconteceram no quarto dia. Cada sessão teve em média 40 participantes, sendo aquelas que ocorreram no terceiro dia (ST-12 à ST-17) as sessões que registraram o maior número de participantes.

A Figura 3 mostra o número de sessões visitadas por congressista. Observa-se uma distribuição exponencial para essa métrica, com a maioria dos participantes

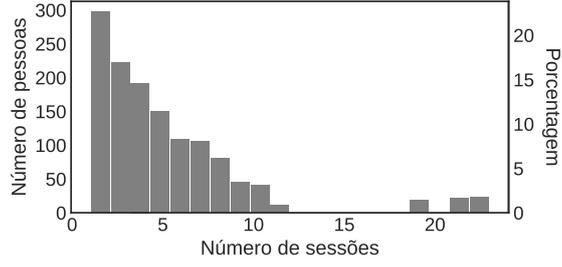


Figura 3. Histograma do número de participantes nas Sessões

($\approx 80\%$) tendo visitado de 1 a 4 sessões e 15% entre 5 e 11 sessões. Cerca de 5% visitaram mais que 19 sessões, o que pode ser explicado, entre outras razões, pelas regras de disputa e de pontuação da gamificação do evento.

4. Modelo de Rede

Os dados de participação nas sessões técnicas foram modelados como um grafo temporal² $G = (V, E, T, \rho)$ com um conjunto de vértices $v_i \in V$ (participantes das sessões técnicas), um conjunto de arestas $e_i \in E$ (contatos), em que $E \subseteq V \times V$, $T \subseteq \mathbb{T}$ em que \mathbb{T} é o domínio temporal do tipo YYYY-MM-DDThh:mm:ss, indica o intervalo de tempo, e $\rho : E \times T \rightarrow \{0, 1\}$ indica se uma aresta e_i existe em um dado tempo. Cada aresta e_i representa a existência do contato entre dois vértices $v_i, v_j \in V \mid v_i \neq v_j$, e possui os atributos ST_i que indica a Sessão Técnica onde o contato aconteceu e $t_i \in T$ que é o horário do último vértice do par (v_i, v_j) a entrar em ST_i . Além disso, cada vértice v_i possui um atributo I_i , representando a instituição ao qual o participante pertence.

Por se tratar de um grafo temporal, a presença de aresta entre dois vértices depende da sessão em que o vértice (participante) se encontra naquele instante de tempo. Na Figura 4 tem-se *snapshots* de um grafo temporal em uma sessão ST_x , exemplificando a adição e remoção de arestas do grafo.

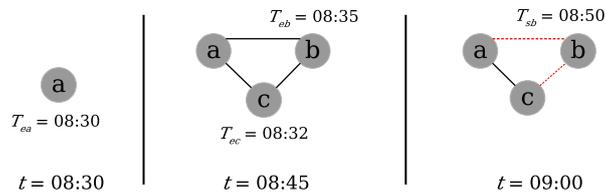


Figura 4. Snapshots de um grafo temporal mostrando a adição e remoção de arestas

Inicialmente o grafo começa vazio, até que no instante T_{ea} identifica-se que o vértice a entrou em ST_x . Alguns minutos depois, identifica-se que nos instantes T_{eb} e T_{ec} , b e c , respectivamente, também entraram em ST_x , e são adicionadas as arestas (a, b) , (a, c) e (c, b) . Dessa forma, a cada instante $t_i \in T$, tem-se um

²Um *grafo temporal* pode ser definido como $\mathcal{G}_\delta(V, A)$, sendo V o conjunto de vértices e A , o conjunto de arestas [Whitbeck et al. 2012]. As arestas são adicionadas ou removidas ao grafo \mathcal{G}_δ do instante t até o tempo $t + \delta$. Existe uma jornada \mathcal{J} de u para v se forem adicionadas arestas ao grafo \mathcal{G}_δ , de tal forma que, o tempo total t' para u alcançar v satisfaça à equação $t' \leq t + \delta$.

grafo completamente conexo, cujos vértices são o conjunto de congressistas presentes na sessão no instante t_i . Com relação à remoção de arestas, como não é possível saber o exato momento em que um vértice saiu de ST_x , considerou-se que, caso seja verificado que o vértice tenha aparecido em outra sessão que ocorre ao mesmo tempo a ST_x , o instante em que o vértice saiu da sessão será igual ao instante em que ele entrou em uma outra sessão. No exemplo da Figura 4, o vértice b permaneceu em ST_x até $T_{sb} = 08 : 50$ (momento em que b entrou em uma sessão simultânea a ST_x), portanto, até esse momento, os vértices a e c possuíam uma aresta com b .

É importante ressaltar que o grafo gerado em cada sessão em um determinado intervalo de tempo é sempre completo, isto é, presume-se que todos em uma mesma sala podem se comunicar entre si. Assume-se, portanto, que uma comunidade é formada pelos congressistas ao final de cada sessão.

5. Caracterização dos Contatos entre Participantes

O objetivo deste trabalho é a caracterização da mobilidade e detecção de comunidades de interesse baseado nas participações das sessões técnicas. Embora não seja possível traçar rotas dos participantes, a análise das características dos contatos é suficiente para definir a iteração e conseqüentemente modelar a rede oportunística que pode ser formada. Portanto, são analisadas as seguintes métricas:

Número de contatos: Contabiliza quantos contatos cada participante teve por dia e durante todo o evento, isto é, o grau de cada vértice ao longo do tempo.

Tempo de contato: Calculado entre participantes e entre instituições. O tempo de contato entre dois vértices a e b ($TC_{a,b}$) em uma sessão é dado por:

$$TC_{a,b} = \begin{cases} T_{ST} - \max(T_{ea}, T_{eb}) & | T_S = 0 \\ T_S - \max(T_{ea}, T_{eb}) & | T_S > 0 \end{cases} \quad (1)$$

onde T_{ST} é a hora que a sessão termina, T_{ea} e T_{eb} são os instantes em que a e b entraram na sessão (respectivamente), $\max(T_{ea}, T_{eb})$ corresponde à entrada mais tardia entre a e b (*i.e.*, quem entrou por último na sessão) e T_S é o instante de saída mais cedo entre a e b (*i.e.*, quem saiu antes da sessão). O tempo de contato entre instituições utilizou a instituição dos participantes como identificador dos vértices.

Tempo entre contatos: Cálculo do tempo que um par de participantes ou instituições demoram para se reencontrarem. Como há a possibilidade de cada participante se conectar com outro em dias diferentes, optamos por contabilizar os tempos de reencontro apenas dos contatos que ocorreram no mesmo dia. Assim, o tempo de reencontro entre dois congressistas pode ser definido pela equação abaixo:

$$TR_{a,b} = T_{ST_{prox}} - T_{ST_{ant}} \quad (2)$$

onde $T_{ST_{ant}}$ e $T_{ST_{prox}}$ são dois pares de sessões em que a e b se encontraram e o encontro ocorrido em $T_{ST_{prox}}$ se deu imediatamente após $T_{ST_{ant}}$. Como a e b podem ter entrado em contato várias vezes no mesmo dia, no final dos cálculos de tempo de reencontro, cada par de participante pode ter um ou mais tempos de reencontro.

Para obtenção desses resultados, criou-se a seguinte estrutura de dados: $\{a : \{(b, ST_x) : [TC, \max(T_{ea}, T_{eb}), S]\}\}$ em que a e b são dois participantes que estavam simultaneamente na sessão ST_x , $TC_{a,b}$ é o tempo de contato entre a e b obtido da Equação 1 e $S \rightarrow \{0, 1\}$ indica se um dos dois participantes foi visto em uma das outras STs simultâneas a ST_x . Dessa forma, o grafo temporal é construído na ordem em que aconteceram as STs no evento.

5.1. Número de Contatos

Para entender como as STs poderiam influenciar na contabilização dos contatos, a Figura 5 mostra a função de distribuição cumulativa complementar (CCDF, na sigla em inglês) do número de contatos apenas nas STs (Fig. 5(a)-b) e considerando os *Workshops* (Fig. 5(c)-d)). Apenas 5% dos participantes realizaram mais de 200 contatos durante o evento, enquanto 90% dos participantes realizaram até 50 contatos. A discrepância entre esses números pode ter sido causada por um pequeno grupo de participantes, que como visto na Figura 3, visitaram acima de 17 sessões. Estes participantes podem ter sido induzidos pela gamificação a visitarem as STs.

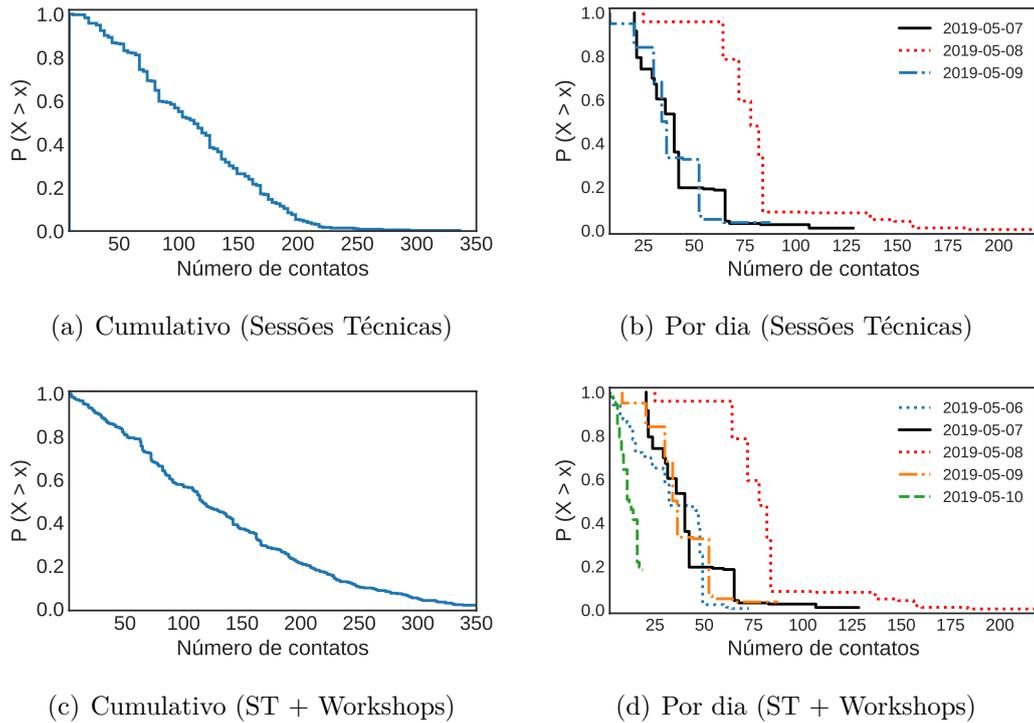


Figura 5. Distribuição cumulativa complementar (ccdf) do número de contatos

A Figura 5(b) apresenta a CCDF do número de contatos que ocorreram durante cada dia do evento. Embora o evento tenha durado 5 dias, as sessões técnicas ocorreram apenas em 3 deles. O segundo (2019-05-07) e o quarto (2019-05-09) dia do evento apresentaram os menores números de contatos, sendo que no segundo dia, $\approx 80\%$ dos participantes realizaram até 24 contatos, enquanto que no quarto dia, 83% dos participantes tiveram até os mesmos 24 contatos. Enquanto isso, no terceiro dia (2019-05-08), mais de 95% dos participantes realizaram até 50 contatos. Como mostra a Figura 2, as sessões que foram visitadas pelo maior número de

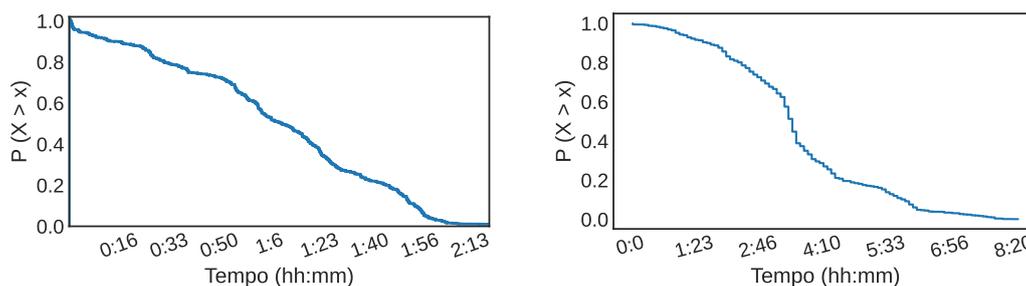
peças ocorreram no terceiro dia do evento, justificando assim, as maiores chances de ocorrerem contatos nesse dia.

A CCDF do número de contatos durante todo o evento ao se considerar, além das Sessões Técnicas, os *Workshops*, é vista na Figura 5(c) a. Obviamente o número de contatos entre os participantes aumentaria com um maior número de sessões observadas, entretanto, é interessante notar que a probabilidade de ocorrerem contatos superiores a 200 é cerca de 10% maior na distribuição da Figura 5(c)

Quando a distribuição do número de contatos é observada por dia do evento, considerando *Workshops* e STs (Fig. 5(d)), há uma discrepância entre o terceiro e último dia do evento. Essa CCDF mostra que, no primeiro dia do evento, $\approx 70\%$ dos participantes realizaram até 20 contatos, com apenas 2% dos congressistas realizando mais do que 50 contatos. Observa-se que a participação nos *Workshops* pouco influenciou nos números de contatos já apresentados antes. A principal causa para isso é que os *Workshops* aconteceram em dias distintos das Sessões Técnicas. Por exemplo, no terceiro dia do evento, 95% dos participantes realizaram até 50 contatos. Já no quarto dia, 84% dos participantes realizaram até 24 contatos, enquanto 32% apresentaram mais do que 50 contatos. No último dia do evento, notou-se um menor número de contatos entre os participantes, sendo que 64% dos congressistas tiveram até 10 contatos, enquanto 18% tiveram 18 contatos, no máximo.

5.2. Tempo de contato

Esta seção apresenta as CCDFs dos tempos de contato entre participantes e entre suas respectivas instituições. Na Figura 6(a) é mostrado a CCDF da duração total dos contatos de cada participante. O tempo de contato entre $\approx 88\%$ dos participantes foi de até 16min. Como pode ser observado na Figura 6(a), um número pequeno de pessoas ($\approx 1\%$) tiveram contatos que duraram acima de 1h e 56min.



(a) Entre participantes

(b) Entre reencontros de participantes

Figura 6. Função distribuição cumulativa complementar (ccdf) do tempo de contato

A Figura 6(b) apresenta a distribuição do tempo entre contato dos participantes, isto é, o tempo que demora para que dois participantes se reencontrem. Como há a possibilidade de reencontros tanto no mesmo dia quanto em dias diferentes, optou-se por calcular a distribuição do tempo entre contatos apenas no mesmo dia. Observa-se que $\approx 60\%$ dos reencontros acontecerem em um intervalo menor do que três horas e vinte e um minutos. Por outro lado, $\approx 15\%$ dos participantes têm tempos entre contato maiores do que 5 horas.

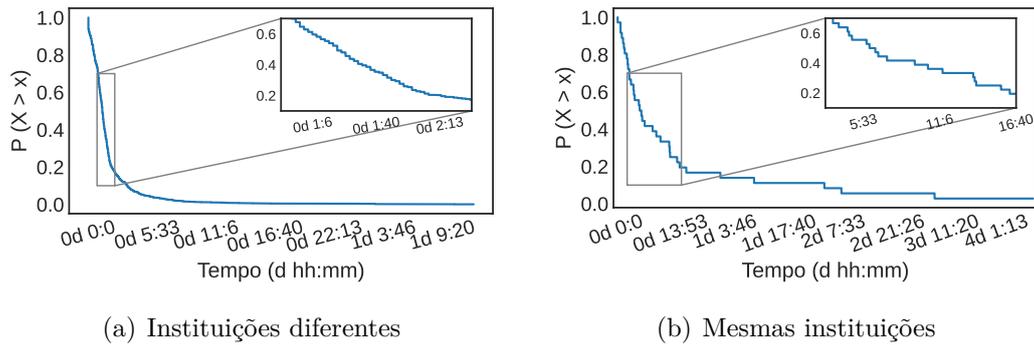


Figura 7. CCDF do tempo de contato entre instituições

Na Figura 7(a) tem-se a distribuição do tempo total de contatos ocorridos entre instituições diferentes. É possível notar, observando a sub figura destacada, que 60% dos participantes tiveram até uma hora de contato com outros participantes. Em contrapartida, apenas 1% dos participantes tiveram tempos de contato acima de 2 horas e 30min. Os tempos de contato maiores que esse valor podem ter sido causados pelos congressistas que visitaram quase todas as sessões do evento.

A CCDF do tempo total de contato entre as mesmas instituições pode ser vista na Figura 7(b). Observa-se que pessoas da mesma instituição passaram um tempo considerável juntas, com cerca de 60% dos participantes tendo até 4 horas de contato. Por outro lado, 2% de participantes da mesma instituição tiveram tempo de contato de até 16 horas e 1% tiveram tempos de contato acima de três dias. Nesses casos, participantes da mesma instituição visitaram as mesmas STs. Em ambos os casos, o tempo de contato apresentou uma distribuição cauda longa.

5.3. Detecção de Comunidades

Esta seção discute como as comunidades são formadas por tópicos de interesse. Na literatura, existem algoritmos conhecidos para identificação de comunidades [Palla et al. 2005, Nguyen et al. 2011, Xu et al. 2013] que podem apresentar desempenhos distintos dependendo das características do grafo que serão utilizadas para identificação de comunidades. A partir da estrutura de dados criada com as informações do evento, entretanto, é possível a identificação das comunidades considerando o interesse do congressista sem a necessidade de se utilizar esses algoritmos.

Inicialmente verificou-se as interseções dos contatos que ocorreram nas sessões técnicas, incluindo as sessões que ocorreram simultaneamente. Embora uma pessoa não possa estar em duas sessões concorrentes, consideramos que, se fosse identificada a presença de duas pessoas em uma sessão, o contato entre elas será contabilizado.

A Figura 8(a) apresenta a detecção de comunidades considerando o interesse dos participantes. No eixo X tem-se o número de sessões visitadas pelas mesmas pessoas e no eixo Y o número de pessoas que visitaram juntas as mesmas sessões. É possível verificar que, quando considera-se a interseção de contatos que ocorreram em 2 sessões, tem-se a possibilidade de até 30 contatos estarem nas 2 sessões consideradas. Quando observa-se interseções entre 3 sessões nota-se que o número de contatos entre elas se reduz pela metade. Por fim, como não há interseção de contatos em mais do que 4 sessões, tem-se que este é o maior número de interseções

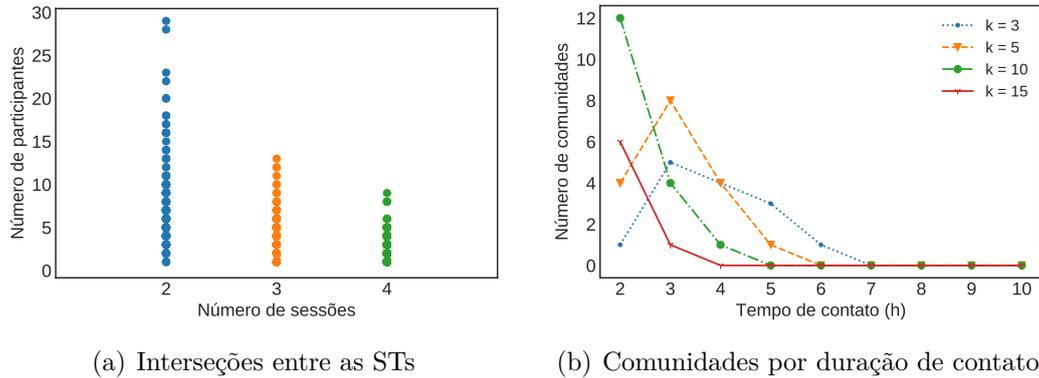


Figura 8. Detecção de Comunidades temporais baseada nas interseções das STs visitadas e por tempo de contato

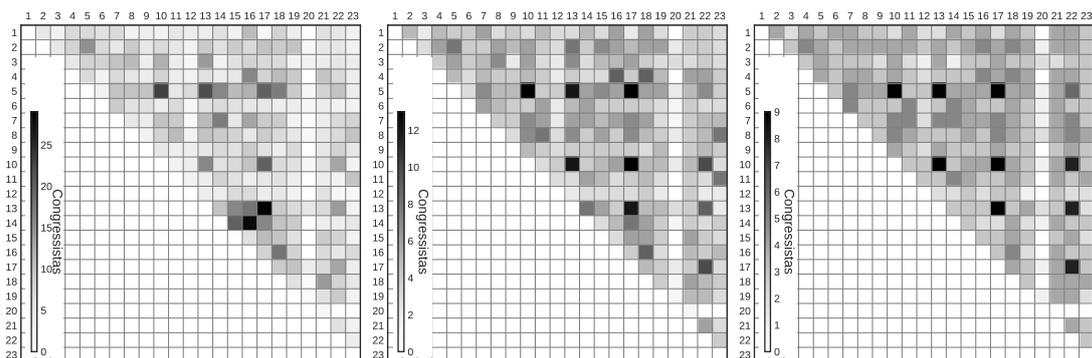
possível, com os maiores clusters sendo de tamanho 9.

Em contrapartida, para compararmos os resultados da identificação de comunidades considerando o tempo de contato entre os participantes, utilizamos o algoritmo de [Palla et al. 2005], permitindo que um mesmo vértice faça parte de uma ou mais comunidades. Em nosso cenário isso é fundamental, tendo em vista que um participante pode fazer parte de comunidades distintas durante o evento. No algoritmo proposto, os autores definem uma comunidade como sendo a união de todos os k -cliques (subgrafos completos de tamanho k) que são alcançáveis entre si a partir de k -cliques adjacentes, ou seja, compartilham $k - 1$ vértices.

Na Figura 8(b) tem-se o número de comunidades por duração de contato entre cada participante. No grafo gerado para a identificação das comunidades, consideramos que há uma aresta entre duas pessoas se elas possuem um tempo de contato maior ou igual a duas horas (maior tempo de duração de uma ST). Além disso, definimos como 3 o número mínimo de pessoas para a formação de comunidades ($k = 3$). Dessa forma, é possível observar na Figura 8(b) que o maior número de comunidades com 10 e 15 participantes é, respectivamente, 12 e 6, e ocorrem entre pessoas que apresentaram um tempo de contato de pelo menos 2 horas. Ressalta-se que quando considera-se tempos de contato superiores a 2 horas foram detectadas apenas as comunidades com 3 participantes, apresentando quantidade de comunidades baixas quando considera-se tempos de contato acima de 6 horas. Outro fato interessante é que apesar de existirem pessoas com um tempo de contato de até 10 horas, não foi possível identificar comunidades com 3 participantes ou mais para tempos de contato superiores a 6 horas.

Para analisar a mobilidade dos congressistas entre cada uma das interseções da Fig. 8, a Figura 9 apresenta os grafos de transições de congressistas entre as STs. O grafo é representado como uma matriz $z_{l,c}$ de adjacências, onde os vértices (Linhas e Colunas) são as sessões técnicas e as arestas têm como peso o número de congressistas que transitaram entre as sessões l e c , sendo que cores mais escuras representam um número maior de participantes na transição.

Verificou-se que nas interseções de 2 STs (Figura 9(a)), as STs (14,16) e (13,17) têm as maiores transições (acima de 25 congressistas). Nas interseções de 3



(a) Interseções entre 2 STs (b) Interseções entre 3 STs (c) Interseções entre 4 STs

Figura 9. Grafos de mobilidade de congressistas por tópicos de interesse representados por matrizes de adjacências. As arestas têm como peso as interseções dos congressistas que visitaram as STs. Índice Linha x Coluna indica o número da ST.

e 4 STs Figuras (9(b) e 9(c)) as transições com mais participantes são (5,10), (5,13), (5,17), (10,13), (10,17) e (13,17), com 12 e 9 participantes, respectivamente. Na programação do evento verifica-se que as STs 14 (Cidades Inteligentes e Big Data) e 16 (Segurança em IoT) bem como as STs 5 (Aprendizado de Máquina na Medição e Monitoração de Redes) e 10 (Engenharia e Controle de Tráfego) tratam de temas similares, indicando que as pessoas nessas sessões têm interesses comuns. Exceção ocorre na transição entre as STs 13 e 17, que tratam de temas distintos e muitos congressistas fizeram essa transição.

6. Implicações do modelo em redes oportunísticas

A eficiência de uma rede oportunística está diretamente relacionada às características dos nós que a compõe, tais como tempo de contato, entre contatos, o número de contatos feitos e as formação de comunidades. Nesse sentido, com relação ao número de contatos, mostramos que há a probabilidade de $\approx 40\%$ dos participantes se encontrarem com pelo menos 32% de todos os congressistas. Alguns *outliers* encontraram acima de 43% dos congressistas e, possivelmente, seriam bons candidatos na disseminação de mensagens nessa rede, por exemplo.

Do ponto de vista de redes móveis, considerando que 88% dos contatos tiveram duração inferior a 16 minutos, seria tempo suficiente para transmitir de ≈ 220 MBytes à 2 Mbps (Bluetooth 5.0) ou 12GB à 100Mbps (IEEE 802.11n), teoricamente. Além disso, considerando que 95% dos reencontros foram inferiores a 1h e 23min, em um período de 8 horas (duração de um dia do evento) os mesmos nós poderiam se encontrar pelos menos 5 vezes durante cada dia do evento.

Utilizando as instituições dos participantes para definir a existência de uma relação social prévia entre eles, identificou-se que $\approx 60\%$ dos congressistas tiveram até 1 hora de contato com pessoas que não faziam parte da sua rede social, i.e., em média 20min por dia de evento. Em teoria, seria possível transmitir 345 MBytes à 2 Mbps (Bluetooth 5.0) ou ≈ 17 GB à 100Mbps (IEEE 802.11n).

Por fim, a detecção de comunidades baseada em tópicos de interesses permite caracterizar as transições dos participantes entre esses tópicos. Isto pode comple-

mentar análises em ambientes acadêmicos cujos contatos em redes oportunísticas é influenciada pelos tópicos de interesse. Nestes casos, os interesses dos participantes podem ser levados em consideração, juntamente com o vínculo social, para definir as comunidades durante um evento.

7. Considerações Finais

Nesse trabalho apresentamos uma análise de comunidades temporais com base em interesses, utilizando como estudo de caso dados coletados durante a 37a edição do SBRC. Modelamos os dados coletados em um grafo temporal, e a partir desse grafo foi possível analisar o número de contatos entre os participantes, o tempo de contato entre participantes e instituições, e o intervalo de tempo de entre os contatos.

Propomos um modelo de detecção de comunidades baseado nas sessões técnicas visitadas pelos participantes e constatamos que a tendência é de que menos comunidades são formadas à medida em que se aumentava o tempo de contato entre os congressistas. Além disso, o tempo de contato aumenta devido aos tópicos de interesse em comum, permitindo detectar as comunidades baseado nas STs visitadas.

Um dos principais desafios para a realização deste trabalho foi a definição do tempo que um participante permaneceu em uma sessão. Foi preciso que identificássemos o horário que uma pessoa entrou em outras sessões para supor quanto tempo ela ficou na sessão que se encontrava atualmente. Além disso, os resultados obtidos podem ter sido influenciados pela gamificação do evento. Por exemplo, não há como saber se uma pessoa participou de uma determinada sessão por interesse genuíno no tópico ou porque desejava aumentar sua pontuação.

Como trabalhos futuros, pretendemos criar um modelo sintético de mobilidade que represente o comportamento real das pessoas em eventos científicos, ou seja, que considere os interesses dos participantes em um evento. A partir deste modelo, pretende-se avaliar protocolos e aplicações em redes móveis e oportunísticas.

Referências

- Akabane, A. T., Immich, R., Madeira, E., and Villas, L. (2019). Aplicando redes sociais veiculares para aprimorar o gerenciamento da mobilidade urbana. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1070–1083. SBC.
- Appel, A. P., Cunha, R. L. F., Aggarwal, C. C., and Terakado, M. M. (2019). Temporally evolving community detection and prediction in content-centric networks. In *Machine Learning and Knowledge Discovery in Databases*, pages 3–18, Cham. Springer International Publishing.
- Awuor, F. M., Wang, C.-Y., and Tsai, T.-C. (2018). Motivating content sharing and trustworthiness in mobile social networks. *IEEE Access*, 6:28339–28355.
- Dede, J., Förster, A., Hernández-Orallo, E., Herrera-Tapia, J., Kuladinithi, K., Kuppusamy, V., Manzoni, P., bin Muslim, A., Udugama, A., and Vatandas, Z. (2017). Simulating opportunistic networks: Survey and future directions. *IEEE Communications Surveys & Tutorials*, 20(2):1547–1573.

- Höyhtyä, M., Apilo, O., and Lasanen, M. (2018). Review of latest advances in 3gpp standardization: D2d communication in 5g systems and its energy consumption models. *Future Internet*, 10(1):3.
- Marques-Neto, H. T., Xavier, F. H., Xavier, W. Z., Malab, C. H. S., Ziviani, A., Silveira, L. M., and Almeida, J. M. (2018). Understanding human mobility and workload dynamics due to different large-scale events using mobile phone data. *Journal of Network and Systems Management*, 26(4):1079–1100.
- McGeehan, D., Madria, S., and Lin, D. (2019). Effective social-context based message delivery using chitchat in sparse delay tolerant networks. *Distributed and Parallel Databases*, pages 1–38.
- Mota, V. F., Cunha, F. D., Macedo, D. F., Nogueira, J. M., and Loureiro, A. A. (2014). Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, 48:5–19.
- Nguyen, N. P., Dinh, T. N., Tokala, S., and Thai, M. T. (2011). Overlapping communities in dynamic networks: their detection and mobile applications. In *17th Int'l conference on Mobile computing and networking*, pages 85–96. ACM.
- Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *nature*, 435(7043):814.
- Pietiläinen, A.-K. and Diot, C. (2012). Dissemination in opportunistic social networks: The role of temporal communities. In *Symposium on Mobile Ad Hoc Networking and Computing*, pages 165–174, New York, NY, USA. ACM.
- Trifunovic, S., Kouyoumdjieva, S. T., Distl, B., Pajevic, L., Karlsson, G., and Plattner, B. (2017). A decade of research in opportunistic networks: challenges, relevance, and future directions. *IEEE Communications Magazine*, 55(1):168–173.
- Viard, J., Latapy, M., and Magnien, C. (2015). Revealing contact patterns among high-school students using maximal cliques in link streams. In *Conference on Advances in Social Networks Analysis and Mining*, pages 1517–1522. IEEE.
- Wang, R., Yang, H., Wang, H., and Wu, D. (2016). Social overlapping community-aware neighbor discovery for d2d communications. *IEEE Wireless Communications*, 23(4):28–34.
- Whitbeck, J., Dias de Amorim, M., Conan, V., and Guillaume, J.-L. (2012). Temporal reachability graphs. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 377–388. ACM.
- Wu, J., Chen, Z., and Zhao, M. (2019). Information cache management and data transmission algorithm in opportunistic social networks. *Wireless networks*, 25(6):2977–2988.
- Xu, H., Hu, Y., Wang, Z., Ma, J., and Xiao, W. (2013). Core-based dynamic community detection in mobile social networks. *Entropy*, 15(12):5419–5438.
- Zhao, Y. and Song, W. (2017). Survey on social-aware data dissemination over mobile wireless networks. *IEEE Access*, 5:6049–6059.