

## V Workshop de Dissertações de Mestrado do PPgSI (2018)

### FICHA DE PESQUISA

DADOS GERAIS				
<b>Título do projeto de pesquisa</b>	Caracterização de eventos de exceção e de seus respectivos impactos no sistema de transporte público por ônibus da cidade de São Paulo			
<b>Orientando</b>	Felipe Cordeiro Alves Dias			
<b>Orientador(es)</b>	Daniel de Angelis Cordeiro			
<b>Semestre no curso, na data do workshop</b>	<input type="checkbox"/> 2º semestre	<input type="checkbox"/> 3º semestre	<input checked="" type="checkbox"/> 4º semestre	<input type="checkbox"/> 5º semestre
<b>Qualificação</b>	[X] Qualificação já realizada em: 23/11/2017 <input type="checkbox"/> Realização da qualificação planejada para:			
<b>Defesa</b>	Prazo máximo para depósito: 29/01/2019		Realização da defesa planejada para: 15/10/2018	
<b>Linha e Área de pesquisa</b>	Gestão e Desenvolvimento de Sistemas: <input type="checkbox"/> BD <input checked="" type="checkbox"/> Engenharia de Software <input type="checkbox"/> Gestão de TI <input type="checkbox"/> IHC		Inteligência de Sistemas: <input type="checkbox"/> IA <input type="checkbox"/> Processamento Gráfico <input type="checkbox"/> Reconhecimento de Padrões	
<b>Área de aplicação</b>	<input type="checkbox"/> Ambientes corporativos	<input type="checkbox"/> Educação a distância	<input type="checkbox"/> Língua Natural	<input checked="" type="checkbox"/> Redes Sociais
	<input type="checkbox"/> Bioinformática	<input type="checkbox"/> Governo eletrônico	<input type="checkbox"/> Linguística	<input type="checkbox"/> Robótica
	<input type="checkbox"/> Biometria	<input type="checkbox"/> Internet	<input type="checkbox"/> Processos de Negócio	<input type="checkbox"/> Saúde
	<input type="checkbox"/> Dispositivos móveis	<input type="checkbox"/> Jogos	<input type="checkbox"/> Químioinformática	<input type="checkbox"/> [outro – escrever]
	<input type="checkbox"/> Educação	<input type="checkbox"/> Jogos sérios		
<b>Publicações associadas ao projeto de mestrado</b>	1) 37th IEEE International Symposium on Reliable Distributed Systems - 1st Workshop on the Distributed Smart City (WDSC'2018). Visualizing large datasets: A case study with data of the buses of São Paulo city – “aceita”. 2) A tweets-based methodology to find bus lines impacted by exception events in the São Paulo city – “em elaboração”.			

DESCRIÇÃO DO PROJETO DE PESQUISA	
<b>Contextualização / motivação</b>	Hoje é possível utilizar as Redes Sociais na caracterização de problemas do transporte público. Apesar disso, não há estudos que correlacionam <i>tweets</i> a dados abertos ( <i>AVL - Automatic Vehicles Location</i> e <i>GTFS – Google Transit Feed Specification</i> ) para análise do impacto de eventos de exceção no transporte público, por ônibus, da cidade de São Paulo.
<b>Problema de pesquisa</b>	Como extrair <i>features</i> de <i>tweets</i> provenientes de <i>profiles</i> oficiais sobre eventos de SP, capazes de contribuir com o melhor entendimento de como os eventos da cidade impactam o transporte público de SP, por meio de dados AVL e GTFS; abordando problemas como processamento distribuído e em quase tempo real de grandes volumes de dados, geolocalização, confiabilidade, exploração e análise de dados.
<b>Objetivo geral da pesquisa</b>	Processar em quase tempo real grandes volumes de dados não estruturados, em Português Brasileiro, tendo o <i>Twitter</i> como fonte, visando extrair <i>features</i> relacionadas ao transporte público de São Paulo para correlacioná-las com dados AVL e GTFS, contribuindo no melhor entendimento de como os eventos de exceção da cidade de SP podem impactar a mobilidade urbana.
<b>Trabalhos relacionados</b>	<i>PairFac: Event Analysis through Discriminant Tensor Factorization</i> (WEN; LIN; PELECHRINIS, 2016), sobre o impacto dos ataques terroristas em Paris no uso do transporte público. <i>Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems</i> (CHEN et al., 2016), sobre o impacto de eventos relacionados ao tráfego na demanda por bicicletas, em Nova Iorque e Washington D.C, USA. <i>Visual Exploracion of Changes In Passenger Flows and Tweets on Mega-City Metro Network</i> (ITOH et al., 2016), sobre o impacto dos eventos anormais nas tomadas de decisão dos passageiros do Metrô de Tokyo. Os trabalhos mencionados não estudam o impacto dos eventos de exceção no contexto da cidade de São Paulo; também, não correlacionam <i>tweets</i> com dados AVL e GTFS.
<b>Proposta para solução</b>	(i) Um abordagem para extração de localização de eventos de exceção por meio de Expressão Regular e obtenção de dados de geolocalização, consultando a <i>API</i> do <i>Google Maps Geocoding</i> ; (ii) extração das demais <i>features</i> por meio de <i>Processamento de Linguagem Natural</i> aplicado aos <i>tweets</i> ; (iii) correlação com a base histórica de dados AVL e GTFS da SPTrans para análise dos impactos do evento.
<b>Dados</b>	(i) Dados das contas oficiais no <i>Twitter</i> responsáveis por reportar eventos de exceção da cidade de São Paulo; (ii) dados históricos AVL referentes a movimentação dos ônibus de SP em 2017 e (iii) dados da especificação do

