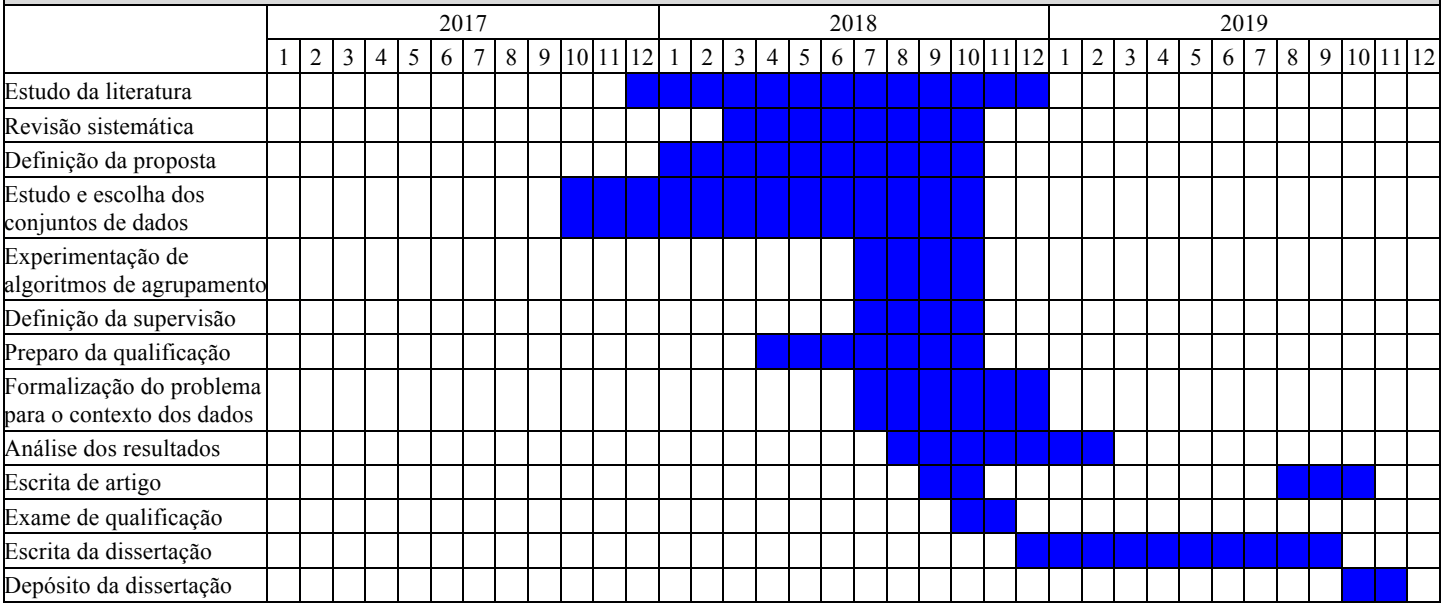
V Workshop de Dissertações de Mestrado do PPgSI (2018)

FICHA DE PESQUISA

DADOS GERAIS				
Título do projeto de pesquisa	Agrupamento interativo aplicado à mineração de processos.			
Orientando	Thais Rodrigues Neubauer			
Orientador(es)	Profa. Dra. Sarajane Marques Peres			
Semestre no curso, na data do workshop	[ ] 2º semestre	[x] 3º semestre	[ ] 4º semestre	[ ] 5º semestre
Qualificação	[ ] Qualificação já realizada em: não aplicável [x] Realização da qualificação planejada para: 30/10/2018			
Defesa	Prazo máximo para depósito: 28/01/2020 Realização da defesa planejada para: 30/10/2019			
Linha e Área de pesquisa	Gestão e Desenvolvimento de Sistemas: [ ] BD [ ] Engenharia de Software [ ] Gestão de TI [ ] IHC	Inteligência de Sistemas: [x] IA [ ] Processamento Gráfico [ ] Reconhecimento de Padrões		
Área de aplicação	[ ] Ambientes corporativos [ ] Bioinformática [ ] Biometria [ ] Dispositivos móveis [ ] Educação	[ ] Educação a distância [ ] Governo eletrônico [ ] Internet [ ] Jogos [ ] Jogos sérios	[ ] Língua Natural [ ] Linguística [x] Processos de Negócio [ ] Quimioinformática	[ ] Redes Sociais [ ] Robótica [ ] Saúde
Publicações associadas ao projeto de mestrado	Não aplicável.			

DESCRIÇÃO DO PROJETO DE PESQUISA	
Contextualização / motivação	Os modelos de processos são ferramentas essenciais para a gestão de negócios, mas é comum que as organizações não os formalizem e, por isso, muitas vezes desconheçam seus processos. A área de mineração de processos extrai conhecimento dos *logs* de eventos de diferentes fases do processo de negócio (AALST, 2011), desempenhando, portanto, um papel central e estratégico por permitir que os processos sejam automaticamente analisados. Em Maita *et al.* (2017), foi realizado um mapeamento sistemático da literatura que mostra como as áreas de mineração de dados e de processos estão inter-relacionadas e concluíram que a tarefa de agrupamento que se origina da primeira está entre as três tarefas mais utilizadas na segunda. O interesse nessa tarefa pode ser explicado por sua natureza descritiva que permite a identificação de padrões existentes nos dados, além de ela não requerer informação prévia sobre esses padrões (SILVA; PERES; BOSCARIOLLI, 2016).
Problema de pesquisa	A tarefa de agrupamento aplicada à mineração de processos gera informação que melhora a interpretabilidade de suas análises, identificando padrões entre as instâncias de processo. Contudo, um potencial problema inerente à essência não supervisionada dessa tarefa é a necessidade de se fazer suposições para resolver questões relacionadas à função de similaridade escolhida, como por exemplo a questão de ambiguidade. Dado que suposições apresentam altas chances de não serem verificadas no contexto real, a qualidade dos resultados do agrupamento pode ser afetada negativamente. Uma opção de resolução para isso é substituir suposições pelo conhecimento de um especialista, o que é chamado de *agrupamento interativo*, e até onde foi possível verificar, não há estudos que explorem o agrupamento interativo no contexto de mineração de processos.
Objetivo geral da pesquisa	O objetivo é a aplicação do agrupamento interativo no contexto de mineração de processos para verificar como a intervenção do usuário afeta a qualidade dos resultados obtidos na tarefa de agrupamento em relação à melhor definição de grupos, usando os índices de validação internos e externos (i) *Jaccard Coefficient*; (ii) *Índice Dunn*; (iii) *Índice Silhouette*; (iv) *NMI*; e (v) *V measure* para medir a qualidade das partições.
Trabalhos relacionados	Os trabalhos *Interactive document clustering with feature supervision through reweighting* (Hu, Milios e Blustein), *Active semi-supervised affinity propagation clustering algorithm based on pair-wise constraints* (Qi, Yu e Min) e *A cluster-level semi-supervision model for interactive clustering* (Dubey, Bhattacharya e Godbole) apresentam sua proposta de aplicação de certa forma de agrupamento interativo e seus resultados.
Proposta para solução	Geração de conjunto de execuções de agrupamento em conjuntos de dados sintéticos e reais da área de mineração de processos, utilizando algoritmos básicos existentes para essa tarefa e diferentes formas de inclusão de supervisão de especialista para alcançar melhores resultados sem que a supervisão seja custosa.

Dados	Os dados utilizados são *logs* de eventos de diferentes fases de um processo de negócio, buscando, através da tarefa de agrupamento, identificar padrões que tragam conhecimento sobre os modelos do processo.
Validação	Avaliação da qualidade dos resultados (modelos) obtidos na tarefa de agrupamento por meio da aplicação de índices de validação internos, *Índice Dunn* e *Índice Silhouette*, e externos, *Normalized Mutual Information (NMI)* e *V measure*, além da comparação entre modelos, por meio da métrica *Jaccard Coefficient*. Com essas métricas, é possível ainda avaliar a intervenção do usuário por meio da comparação entre os modelos gerados com e sem ela.
Limitações	A primeira limitação identificada é a falta de controle e conhecimento da geração e manutenção dos *logs* estudados. Além disso, a avaliação da intervenção do usuário na qualidade dos grupos obtidos podem não refletir aspectos subjetivos como o alinhamento com a expectativa dele. Por fim, medir o esforço do usuário é uma tarefa também subjetiva e, portanto, o modo adotado pode não refletir completamente diferentes conceitos de esforço.
Resultados esperados	Contribuições científicas: Conjunto de práticas para a aplicação de agrupamento interativo à área mineração de processos, incluindo avaliação do esforço do usuário e de sua intervenção na qualidade do resultado obtido.
	Contribuições tecnológicas: Framework de aplicação de agrupamento interativo a mineração de processos.

MÉTODO DE PESQUISA				
Gênero	[ ] Pesquisa teórica	[ ] Pesquisa prática	[x] Pesquisa empírica	[ ] Pesquisa metodológica
Natureza	[ ] Pesquisa básica/pura	[x] Pesquisa aplicada		
Objetivo	[x] Pesquisa exploratória	[ ] Pesquisa descritiva	[ ] Pesquisa explicativa	
Abordagem	[ ] Pesquisa quantitativa	[ ] Pesquisa qualitativa	[x] Pesquisa mista (quali-quanti)	
Procedimento técnico principal	[x] Pesquisa experimental [ ] Pesquisa bibliográfica [ ] Pesquisa documental [ ] Pesquisa *ex-post-facto* [ ] Pesquisa de levantamento	[ ] Pesquisa com *survey* [ ] Estudo de caso [ ] Pesquisa participante [ ] Pesquisa-ação [ ] Pesquisa etnográfica	[ ] Pesquisa netnográfica [ ] Teoria fundamentada em dados (*Grounded theory*) [ ] Ciência do projeto (*Design science*)	
Técnica(s) / Instrumento(s) de coleta de dados	[ ] Medição [ ] Questionário [ ] Entrevista [ ] Grupos focais	[ ] Formulário [ ] *Benchmark*	[ ] Observação (direta / participante) [ ] Diário de campo / notas de campo [ ] Análise documental (ou de artefatos) [x] Dados de pesquisa anterior	
Técnica(s) de análise de dados	Análise quantitativa: [x] Estatística descritiva [ ] Estatística inferencial [x] Análise multivariada		Análise qualitativa: [x] Análise de conteúdo [ ] Análise do discurso	

CRONOGRAMA