**Nome Completo dos Alunos: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Tarefa Aulas 5 e 6 – Exploração de  
*Cluster Analysis* e SNA em R**

**Instruções**

Esta tarefa deverá ser feita em grupos de até 5 alunos.

Esta tarefa deverá ser respondida e encaminhada ao professor via ECLASS **até o dia 10/Novembro, 23h59**.

Utilizem ESTE DOCUMENTO (BDADS\_2S2023\_Tarefa Aulas 5 e 6.DOCX) como modelo para o documento da tarefa (Times New Roman 12, espaçamento simples entre linhas e antes e depois dos parágrafos, limites de margem conforme este documento).

A folha de respostas deverá ter o nome do(s) aluno(s), o cabeçalho de cada questão e as respostas (*retirem este trecho de instruções da versão final*). O documento da tarefa deverá ter pelo menos 3 páginas. Se quiserem colocar figuras para ilustrar as respostas, elas devem ser colocadas no final (com referência no texto) e não comprometer o limite de páginas para o texto das respostas.

**Questões Base**

* Explore as rotinas **Exemplo Rede One Mode.R** e **Exemplo Rede Two Mode.R** . Rode os códigos na plataforma R utilizando como base as tabelas Rede One Mode\_Tarefa Aulas 5 e 6.xlsx e Rede Two Mode\_Tarefa Aulas 5 e 6.xlsx . Se quiser, faça pequenas modificações na tabela e veja seus resultados (atenção: não são as mesmas bases trabalhadas em sala).
* Utilizando a Rede One Mode do arquivo Rede One Mode\_Tarefa Aulas 5 e 6.xlsx descreva sua estrutura de componentes, nós, arestas, cliques, diâmetro, distância média e densidade.
* Faça pequenas modificações na tabela e veja seus resultados.
* Utilizando a Rede Two Mode do arquivo Rede Two Mode\_Tarefa Aulas 5 e 6.xlsx descreva sua estrutura de componentes, nós, arestas, cliques, diâmetro, distância média e densidade.
* Inclua outras análises em seu código (usando as extensões **sna**, **network** ou **igraph**) e comente os resultados (seja criativo!).
* Compile as saídas dos códigos (conteúdo das variáveis, gráficos, tabelas) em um documento Word (usando o modelo deste documento) e comente seus resultados (principalmente as medidas de centralidade), análises, potenciais implicações gerenciais, etc, conforme discutido em sala nas Aulas 5 e 6.
* Explore os métodos de análise de agrupamentos Hierárquico Aglomerativo (com o auxílio do dendrograma) e *K-means*. Teste diferentes métodos de ligação (*average, single, complete, centroid, Ward*) e escolha a configuração que achar mais adequada.
* Utilize as bases de dados que estão na tabela Cluster Analysis\_Tarefa Aulas 5 e 6.xlsx (qualquer uma delas – Hábitos Alimentares ou Sanduíches). Se quiser, faça pequenas modificações na tabela e veja seus resultados.
* Utilize as ferramentas Minitab ou R, conforme visto em aula.
* Apresente as saídas das suas análises (tabelas, gráficos, medidas resumo, etc) em um documento Word (usando o modelo deste documento) e comente seus resultados conforme discutido em sala nas Aulas 5 e 6.
* Por que as técnicas de agrupamentos se tornaram mais importantes no contexto do *Big Data Analytics*?
* **Desafio:** Baseado na tabela da Rede Two Mode desta tarefa, faça uma análise de agrupamento (*cluster analysis*) do tipo hierárquico aglomerativo (dendrograma) das pessoas ou dos produtos adquiridos por elas, levando em consideração apenas a estrutura de relações entre elas. Comente como implementou e discuta os resultados, comparando com a rede construída. Utilize a plataforma R e o script de exemplo de uso de *Cluster Analysis* em R.

*Dica: após a seleção dos grupos, desenhe a rede e represente os nós das pessoas (ou produtos) com cores de acordo com o grupo correspondente.*