1. Entender os atributos da base.
2. Entender Silhueta, Ensemble.
3. Consenso é igual ao Ensemble?
4. Ensemble = igual a comitê de agrupamento.

# Comitê de agrupamento

Diversos algoritmos de agrupamento estão sendo desenvolvidos nos últimos anos para vários domínios. Entretanto, nenhum desses é adequado para todos os tipos de aplicações e de dados. O algoritmo de agrupamento tem características específicas, tendo uma variedade de soluções para um mesmo conjunto de dados. (Silva Semaan, p. 77, 2013). Nesta definição, surge o comitê de agrupamento do inglês “*Cluster ensemble*”. É uma combinação de conjunto de partições bases que gera uma única solução consenso (Fadel *et al.*, p. 3, 2014).

O processo de combinação de agrupamento emprega uma função de consenso em um conjunto *n* partições, conhecidas como partições base. Essas partições são criadas de um ou mais algoritmos de agrupamento, aplicados em conjunto de dados. O objetivo é que a partição de consenso tenha melhores resultados que as partições base (Piantoni, p. 29, 2016).

Nesse contexto, a combinação de agrupamento tem diferentes objetivos, como (Silva Semaan, 2013) e (Piantoni, 2016):

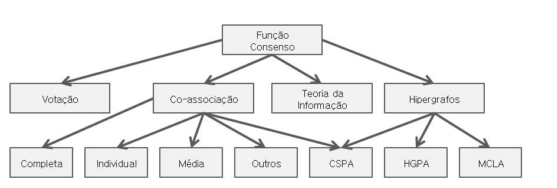
* **Robustez:** Conseguir resultados melhores do que a maioria das partições bases, ou que seja menos sensível a ruídos.
* **Consistência:** Obter uma partição de consenso que tenha concordância com as partições base.
* **Novidade:** Obter partições diferenciadas do que estavam no conjunto de origem.
* **Reaproveitamento de conhecimento:** Usar o conhecimento gerado pelas partições base para gerar a solução de consenso.

# Função de consenso

De acordo com Piantoni (p. 29, 2016), existe uma variedade de função de consenso disponível na literatura, onde cada uma tem métodos específicos para realizar combinações. Segue a definição de algumas:

* **Função baseada em grafo ou hipergrafo**: Usam o método de manipulação de grafo ou hipergrafos para extrair o consenso das partições bases.
* **Função baseada em probabilidade:** Calcula a probabilidade de um objeto estar em uma partição, fundamentando-se nas informações extraídas da partição base.
* **Função baseada em co-associação**: calcula a similaridade entre objetos, através dos números de grupos que estão compartilhados entre eles em todas as partições base.

A figura 1 está disponível no trabalho do Nepomuceno (p. 47, 2008), demonstrando a função de consenso e suas possibilidade.



**Figura 1** estratégia de integração (NEPOMUCENO, p. 47, 2008)

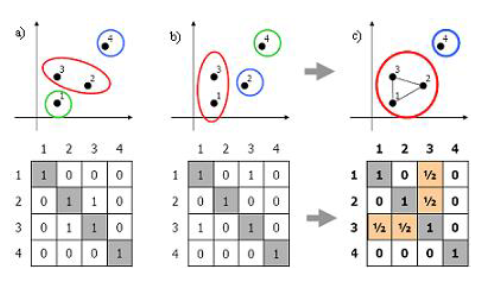
## Função de consenso utilizando matriz de co-associação

É uma técnica que faz agrupamentos por acumulo de evidência. O objetivo é combinar os resultados em uma única partição final, onde cada grupo resultante é uma evidencia independente da organização dos dados. De acordo com (NEPOMUCENO, p. 48, 2008) existe três problemas que precisam ser resolvidos quando é utilizado a abordagem de agrupamento por acumulo de evidencia:

* Como coletar as evidencias?
* Como combinar as evidencias?
* Como extrair uma partição sólida, após a combinação de evidencias?

Respondendo a primeira pergunta, a coleta das evidencias é feita através das partições bases. Para resolver a segunda questão, é utilizada a matriz de co-associação. Nepomuceno (2008) informa que “a similaridade entre os objetos pode, então, ser estimada a partir da quantidade de grupos compartilhados por dois objetos em todas as partições dos dados. Essa similaridade expressa a força de co-associação de cada par de objetos”.

Com os votos da associação, as *N* partições de dados são transformadas em uma matriz de co-associação que tem o tamanho *n* x *n*.A figura 2 (p. 50) demonstra o uso de matrizes de co-associação em partições diferentes, o resultado final é melhor do que as partições separadas.



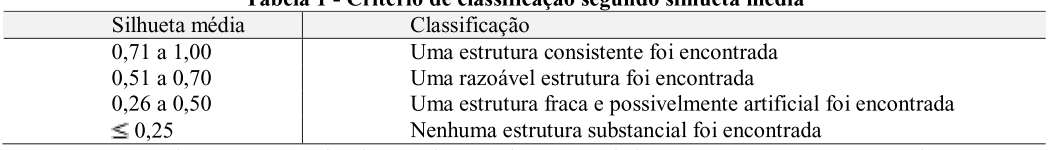
**Finura 2** Representação da primeira e segunda partição para formar a partição final Nepomuceno (p. 50, 2008).

Para resolver o terceiro problema, o Nepomuceno (p. 50, 2008) informa que é possível aplicar um algoritmo de agrupamento a nova matriz de similaridades, até que tenha melhores resultados. Na dissertação de Nepomuceno (p. 50, 2008) foi utilizado um algoritmo hierárquico com ligação média e completa para realizar o agrupamento.

# Silhueta

O algoritmo de agrupamento cria grupos mesmos não existindo uma estrutura natural de grupos nos dados. Por esse motivo, é fundamental validar cada partição gerada. Os índices de validação apresentam meios para validar soluções obtidas de diferentes métodos, não deixando apenas validar, mas ordenar. Uma grande variedade de índices está disponível na literatura. (Fadel *et al.*, p. 5,2015).

Um exemplo para validação das partições no agrupamento é o incide de silhueta. Ele define a qualidade dos grupos com informações sobre a proximidade dos objetos do mesmo grupo e na distância entre objetos do grupo que estão próximo. De acordo com Fadel *et al.*,(p.4, 2014) “Pode ser calculado com qualquer medida de dissimilaridade ou similaridade e retorna valores no intervalo [-1,1], onde valores positivos próximos de 1 indicam que o objeto está bem situado em seu grupo e valores negativos próximos de -1 indicam que o objeto está mais próximo de outro grupo”.



**Tabela 1** - Critério de classificação segundo silhueta média. (Fadel *et al.*, p. 4, 2014)

Fadel, A. *et al.* (2015) ‘PROPOSTA DE MÉTODO PARA SOLUÇÃO DO PROBLEMA DE AGRUPAMENTO AUTOMÁTICO’, *SBPO*. Available at: http://cdsid.org.br/sbpo2015/wp-content/uploads/2015/08/143076.pdf (Accessed: 25 April 2018).

Fadel, A. C. *et al.* (2014) ‘UM ESTUDO DA APLICAÇÃO DE TÉCNICAS DE COMBINAÇÃO DE AGRUPAMENTOS’. Available at: http://pdf.blucher.com.br.s3-sa-east-1.amazonaws.com/marineengineeringproceedings/spolm2014/126207.pdf (Accessed: 25 April 2018).

NEPOMUCENO, V. S. (2008) ‘Algoritmos de agrupamento tradicionais versus sistemas de comitê de agrupamentos: análise de dados de expressão gênica’. Universidade Federal de Pernambuco. Available at: https://repositorio.ufpe.br/handle/123456789/2824 (Accessed: 25 April 2018).

Piantoni, J. (2016) ‘Análise comparativa de técnicas avançadas de agrupamento’. Universidade Federal de São Carlos. Available at: https://repositorio.ufscar.br/handle/ufscar/8252 (Accessed: 25 April 2018).

Silva Semaan, G. (2013) ‘Algoritmos para o Problema de Agrupamento Automático’.