Many non-parametric multi-view clustering models and applications exist:

- A feature partition \( \mathcal{F} = \{A_1, A_2, \ldots\} \) over the data points \( |N| = \{1, \ldots, N\} \) with \( K \) possible features \( |K| = \{1, \ldots, K\} \) is defined as a partition of \( |N| \times |K| \), pairs of natural numbers where the first element denotes the data point label and the second element denotes the feature label. We require all subsets \( A_i \) to have the property that:
  
  \[
  \text{if } (i, j) \in A_i \text{ and } (i', j') \in A_i, \text{ then it must be that } j = j'.
  \]

- A random feature partition \( F \) is a random element in the set of feature partitions, \( \mathcal{F} \), and is said to be exchangeable if
  
  \[
  \sigma_i \times \sigma_j (F) = F
  \]
  
  for all \( \sigma_i \) permutations of \( \{1, 2, \ldots, |N|\} \) and \( \sigma_j \) permutations of \( \{1, 2, \ldots, |K|\} \).

- A random feature partition \( F \) is exchangeable w.r.t. the data points with asymptotic frequencies in order of appearance given by the distribution \( \alpha \). Let \( p(\cdot) \) be defined as
  
  \[
  p\left(\left\{(n_{ij})_{i,j=1}^{K}\right\}\right) = E_{\alpha \sim P}\left[\prod_{j=1}^{K} \prod_{i=1}^{n_{ij}} \prod_{k=1}^{K} \left(1 - \frac{1}{K} \right)^{1 - n_{ij}} P_{ji}ight]\]
  
  for \( \left\{(n_{ij})_{i,j=1}^{K}\right\} \), a sequence of sequences of natural numbers.

- This function generalises the exchangeable random partition probability function and feature allocation probability function.

**Feature Partitions**

**Definition 1.** A feature partition \( \mathcal{F} = \{A_1, A_2, \ldots\} \) over the data points \( |N| = \{1, \ldots, N\} \) with \( K \) possible features \( |K| = \{1, \ldots, K\} \) is defined as a partition of \( |N| \times |K| \), pairs of natural numbers where the first element denotes the data point label and the second element denotes the feature label. We require all subsets \( A_i \) to have the property that:

\[
\text{if } (i, j) \in A_i \text{ and } (i', j') \in A_i, \text{ then it must be that } j = j'.
\]

**Feature Partitions**

**Definition 2.** Given a sequence of probability measures \( \mu_i \) each defined over the interval \([0, 1]\) with disjoint support sets, for every \( j \) generate a sequence of random variables \( X_{ij} \sim \mu_i \) iid for \( i \in \mathbb{N} \). The sequence \((X_{ij})\) defines a random feature partition \( F \) exchangeable in data points by \( F_{ij} = (\omega X_{ij} = X_{i}^\prime(j)) \), the event that \( s \) and \( y \) belong to the same block for \( x = (s, j), y = (i', j') \). If for all \( j \) we have in addition that if \( \mu_i \sim \mu \) then \( F \) is exchangeable in features as well.

The following construction extends Kingman’s paintbox to feature partitions:

**Definition 3.** The underlying directing measure for a feature partition \( \mathcal{F} \) is a factorial paintbox construction with some random measure \( \alpha \) over random measures \( \{\alpha_i\} \).

**Exchangeable Probability Function**

**Theorem 3.** The underlying directing measure for a feature partition is a factorial paintbox construction with some random measure \( \alpha \) over random measures \( \{\alpha_i\} \).

**Impact**

Many models and applications use multi-view clustering...

- ... we identified various multi-view clustering models as equivalent,
- ... we collapsed many multi-view clustering applications into the same class,
- ... we can share algorithmic insights between the various models using the feature partition as their underlying model:
  
  - Explain away differences between the models,
  - Unify inference for the different models (Gibbs sampling used for some and variational inference for others)
  - A clear way of generalisation (using various distributions over the partitions, introducing new dependencies, etc.).

**Future Research**

Introduce dependencies to the factorial paintbox construction:

- Could be depicted as each block having its own paintbox sampled from a distribution conditioned on the block itself.
- Could be used to model the underlying structure of correlated multi-clustering models such as in Doshi-Velez and Ghahramani [2009],
- Corresponds to a special case of the fragmentation chain [Bertoin, 2006].