



# A dependent partition-valued process for multitask clustering and time evolving network modelling

Konstantina Palla, David A. Knowles, Zoubin Ghahramani

(Submitted on 13 Mar 2013)

The fundamental aim of clustering algorithms is to partition data points. We consider tasks where the discovered partition is allowed to vary with some covariate such as space or time. One approach would be to use fragmentation-coagulation processes, but these, being Markov processes, are restricted to linear or tree structured covariate spaces. We define a partition-valued process on an arbitrary covariate space using Gaussian processes and a novel interpretation of the stick breaking construction. By choosing the parameters of the stick breaking construction the process can be given Chinese restaurant process or Pitman Yor process marginals. We use the process to construct a multitask clustering model which partitions datapoints in a similar way across multiple data sources, and a time series model of network data which allows cluster assignments to vary over time. We use Elliptical Slice Sampling for inference and apply our method to defining cancer subtypes based on different types of cellular characteristics, finding regulatory modules from gene expression data from multiple human populations, and discovering time varying community structure in a social network.

Comments: 10 pages, 7 figures, submitted for review

Subjects: **Machine Learning (stat.ML)**

Cite as: [arXiv:1303.3265](#) [stat.ML]

(or [arXiv:1303.3265v1](#) [stat.ML] for this version)

## Submission history

From: Konstantina Palla Miss [[view email](#)]

[v1] Wed, 13 Mar 2013 13:55:20 GMT (539kb,D)

*[Which authors of this paper are endorsers?](#)*

## Download:

- [PDF](#)
- [Other formats](#)

Current browse context:

stat.ML

[< prev](#) | [next >](#)

[new](#) | [recent](#) | [1303](#)

Change to browse by:

[stat](#)

## References & Citations

- [NASA ADS](#)

## Bookmark([what is this?](#))

