Representaciones óptimas de sistemas complejos

Ariel Haimovici, DF

Pablo Balenzuela, DF

Dante Chialvo, UNSAM

Enzo Tagliazucchi, Netherlands institute for Neuroscience, Amsterdam, The Netherlands

Matteo Marsili, ICTP, Trieste, Italia

El problema general: Inferencia en sistemas complejos

$$\hat{S} = \begin{array}{c} \text{variables} \\ 1 \quad \dots \quad N \\ \left(\begin{array}{cccc} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{array} \right) \begin{array}{c} 1 \\ \\ M \end{array} \end{array} \text{ instancias}$$

Por ejemplo:
- Series temporales en registros neuronales
- Series financieras
- Secuencias de amino ácidos en proteínas

$\hat{S}=(s^{(1)},\dots,s^{(M)})$ Queremos inferir el proceso generativo: P(s)

Típicamente: $|\Omega| \sim \exp(N) >> M$

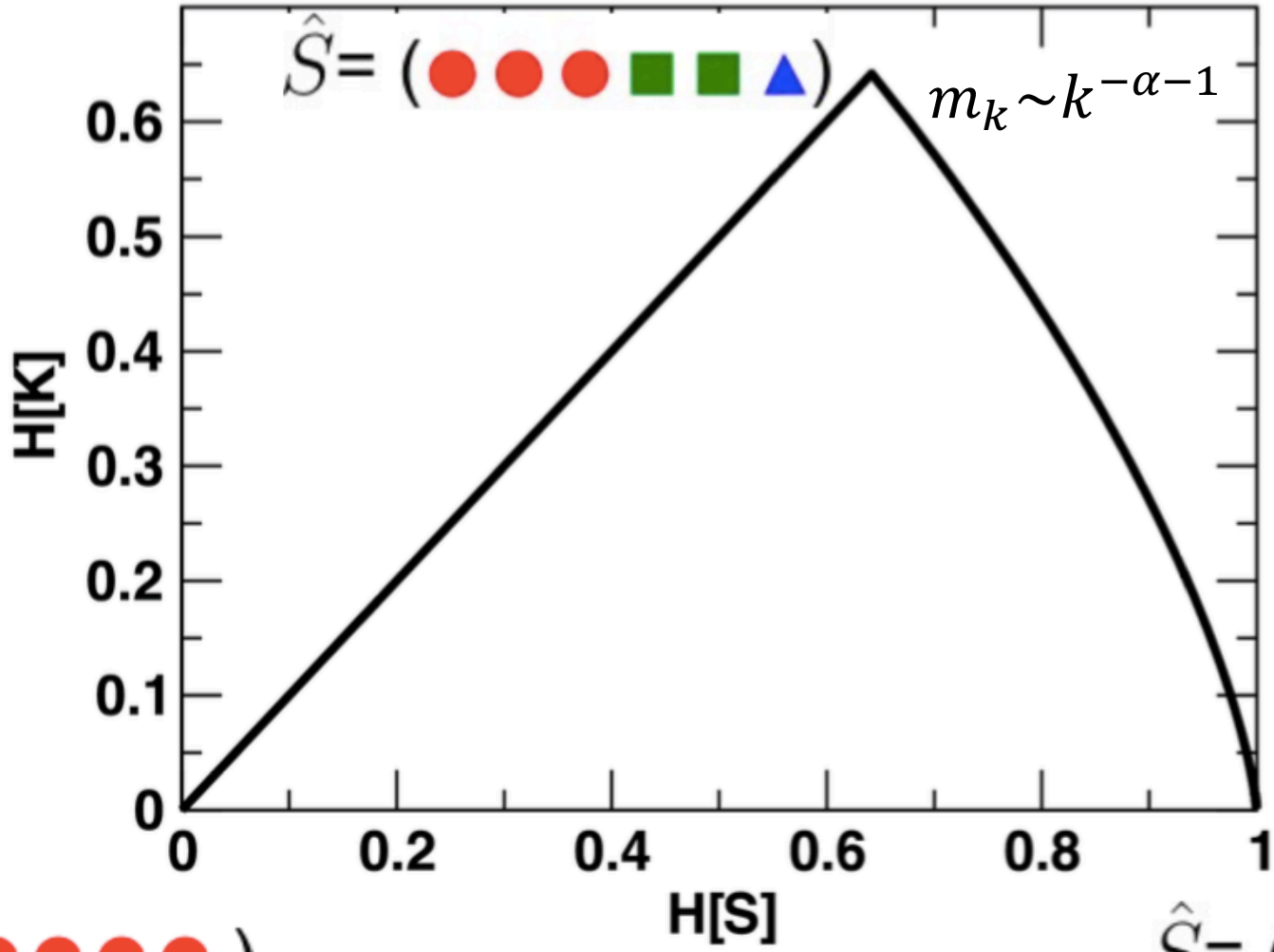
Reducción de dimensionalidad: selección de variables, clustering

Resolución y Relevancia

$\hat{S}=(s^{(1)},\ldots,s^{(M)})$

$$H[S] = -\sum_{s} \frac{k_s}{M} \log(\frac{k_s}{M})$$

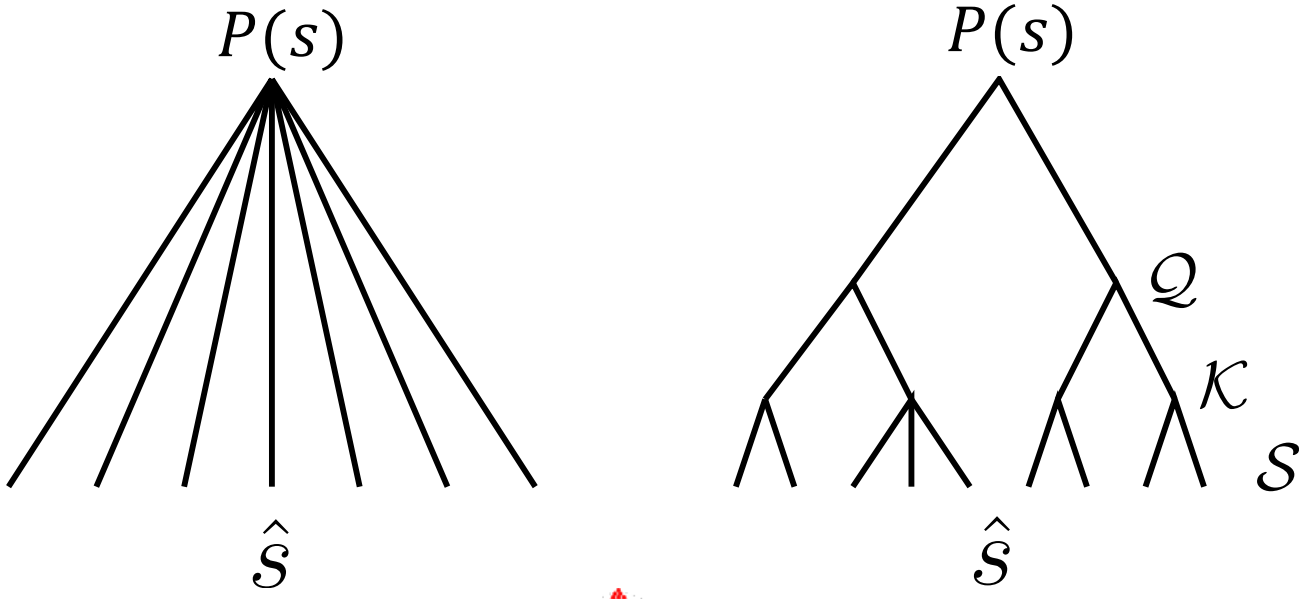
$$H[K] = -\sum_{k} \frac{m_k k}{M} \log(\frac{m_k k}{M})$$

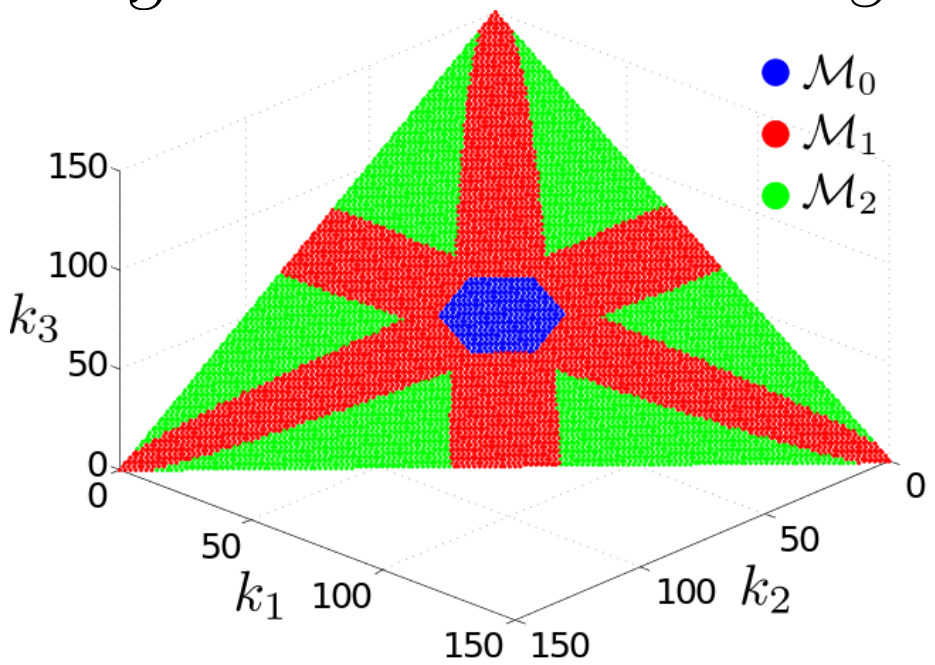


$\hat{S}$= (● ● ● ● ● ●)

$\hat{S}$= (● ■ ▲ ⬟ ★ ▬)

Selección de modelos desde un punto de vista Bayesiano





$$P(\hat{S}|\mathcal{M}) = \prod_s p_s{}^{ks} \rightarrow p_s = \frac{k_s}{M}$$

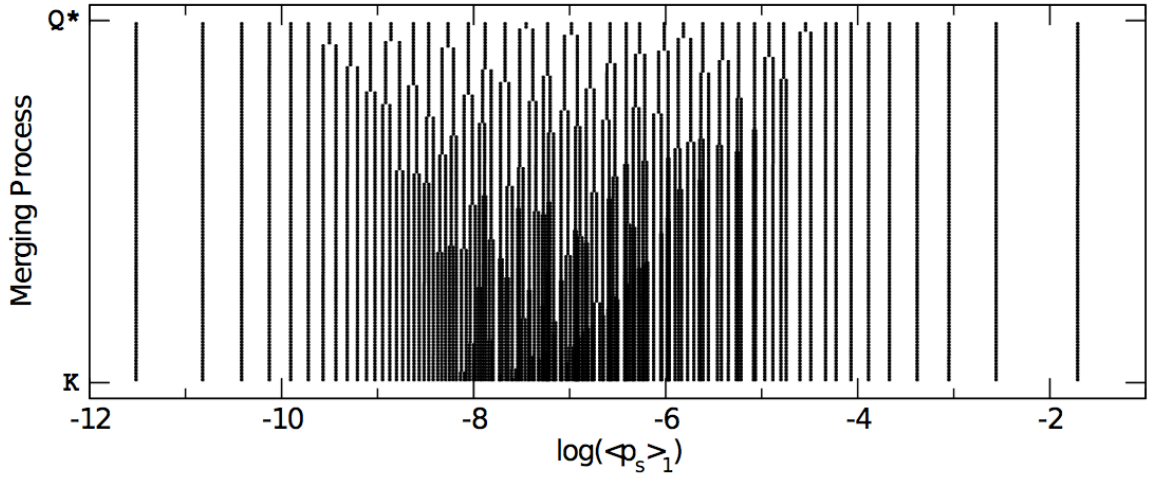
$$P(\mathcal{M}_i|\hat{S}) = \frac{P(\hat{S}|\mathcal{M}_i)P_0(\mathcal{M}_i)}{\sum_j P(\hat{S}|\mathcal{M}_j)P_0(\mathcal{M}_j)}$$

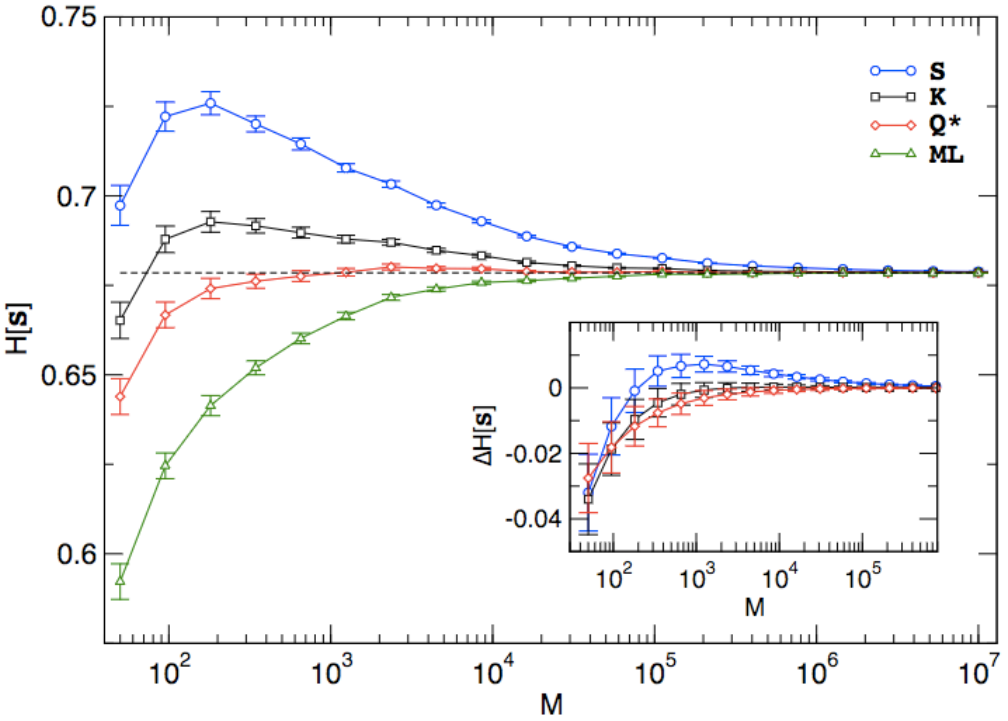
$$\mathcal{M} = [\mathcal{Q}, \vec{\mu}]$$

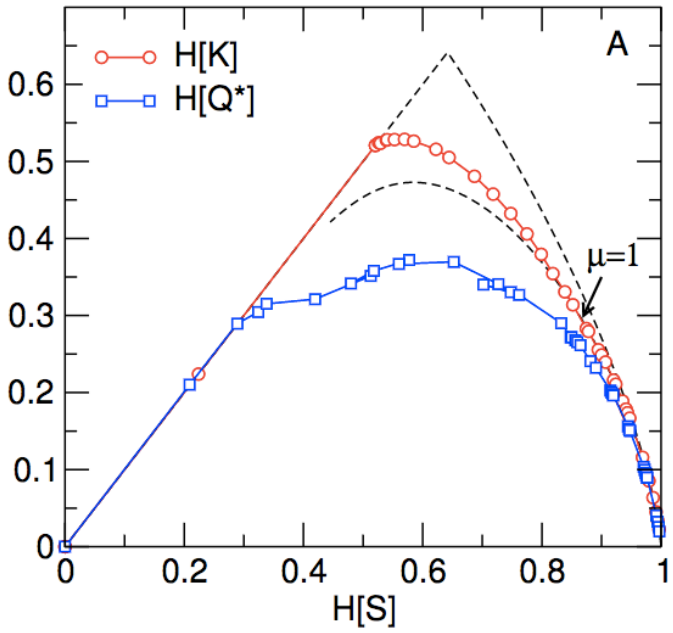
$$P(\mathcal{Q}|\hat{S}) \sim P(\hat{S}|\mathcal{Q}) = \int d\vec{\mu} \prod_q \mu_q{}^{Kq} \; P_0{}^{(\mathcal{Q})}(\vec{\mu})$$

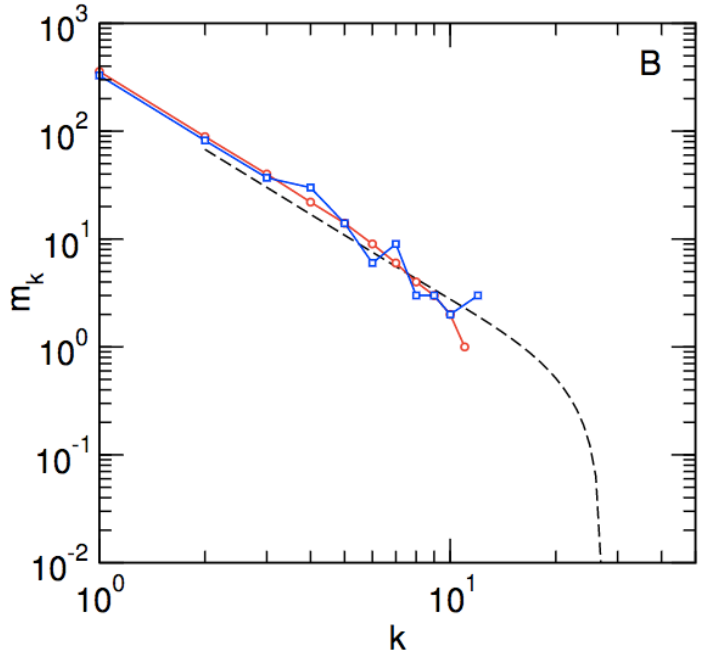
$$K_q = \sum_{s \in Q_q} k_s$$

Las representaciones “óptimas”:









Aplicaciones en curso

- Selección de posiciones relevantes en secuencias de proteínas
- Clustering de estados cerebrales medidos con fMRI

¡Gracias!

Referencias:
Marsili, Mastromatteo, Roudi, Journal of Statistical Mechanics, 2013
Haimovici, Marsili, Journal of Statistical Mechanics, 2015