Quantifying Network Similarity using Graph Cumulants

Gecia Bravo-Hermsdorff [1] Lee M. Gunderson[1] Pierre-André Maugis[2] Carey E. Priebe[3]

[1]Department of Statistical Science, University College London [2]Google Research Zürich [3]Department of Applied Mathematics and Statistics, Johns Hopkins University

TL;DR: Graph cumulants perform better and are more intuitive than the typical subgraph statistics

Using only moments is awkward

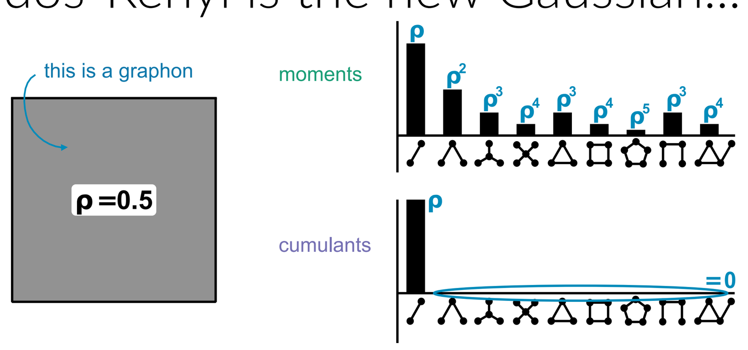
"The length of a human averages 1.7 meters, and their **average squared length** is 2.9 square meters"

Using cumulants is **easier** to understand

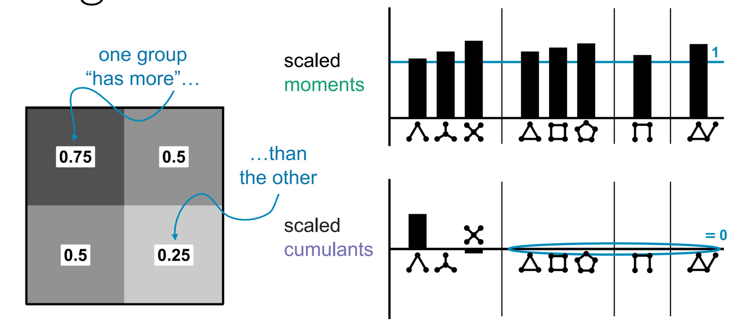
"The length of a human has: a **variance** of 0.01 meters, a **standard deviation** of 0.1 meters, a **relative fluctuation** of 6%"

Graph Cumulants: the Better Subgraph Statistics

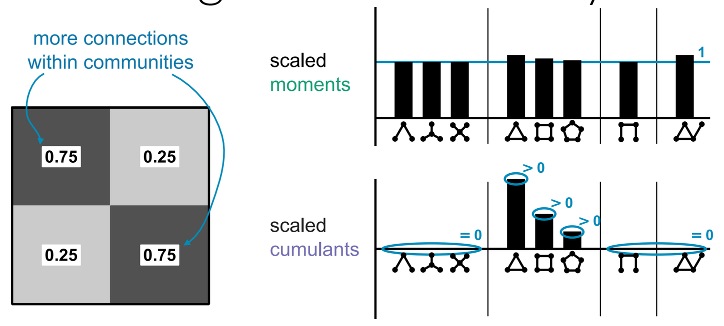
Erdős-Rényi is the new Gaussian...



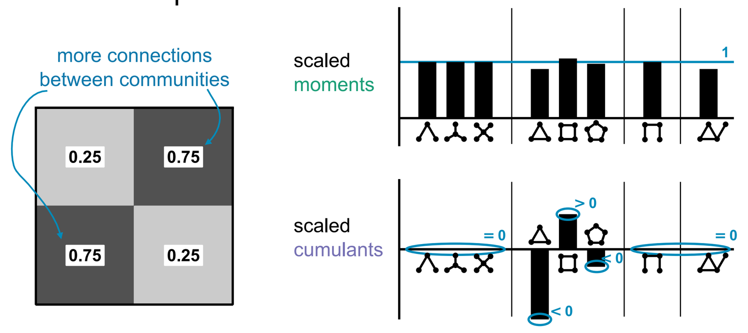
...degrees are encoded in stars...



...clustering is encoded in cycles...

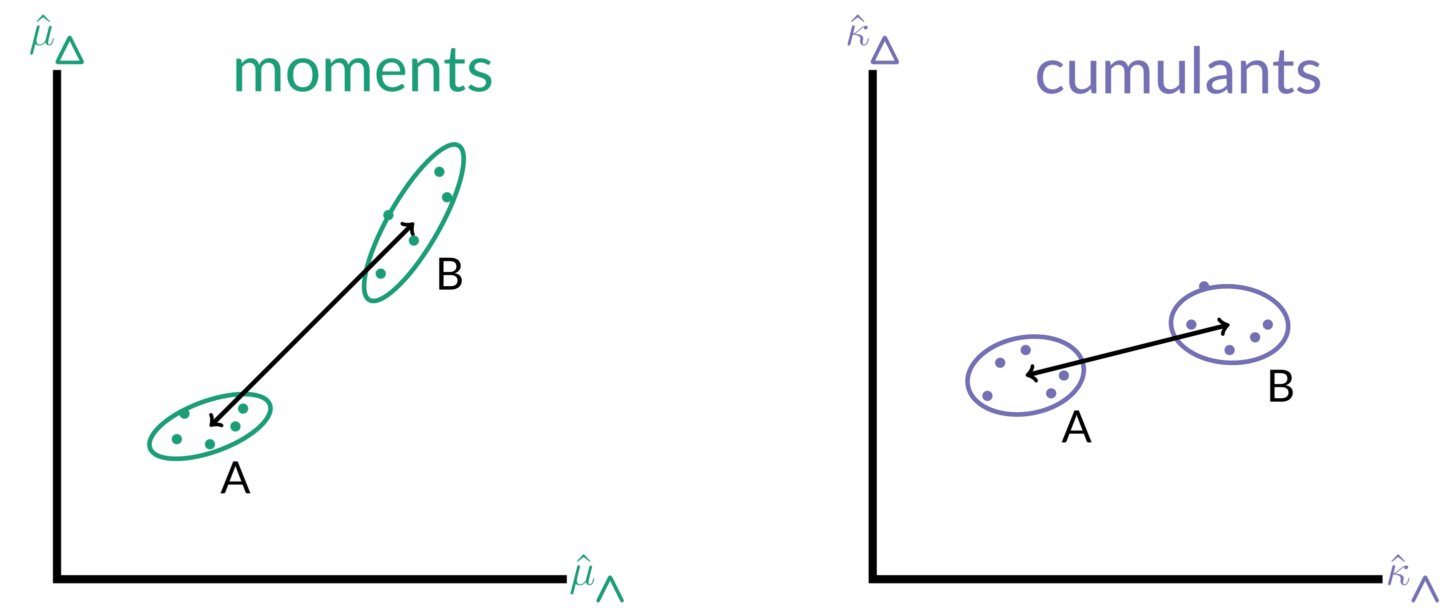


...and bipartite-ness as well!



But are graph cumulants better for testing? (than subgraph densities)

Apples-to-Apples: A Two-Sample Test



$$\underbrace{\hat{d}^2_{\kappa}(\mathsf{A},\mathsf{B})}_{\text{“}Z^2\text{-score”}} = \underbrace{\left(\hat{\kappa}(\mathsf{A}) - \hat{\kappa}(\mathsf{B})\right)^{\top}}_{\text{difference}} \underbrace{\left(\hat{\Sigma}_{\kappa}(\mathsf{A}) + \hat{\Sigma}_{\kappa}(\mathsf{B})\right)^{-1}}_{\text{metric}} \underbrace{\left(\hat{\kappa}(\mathsf{A}) - \hat{\kappa}(\mathsf{B})\right)}_{\text{difference}}$$

$$\hat{d}^2_{\mu}(\mathsf{A},\mathsf{B}) = \left(\hat{\mu}(\mathsf{A}) - \hat{\mu}(\mathsf{B})\right)^{\top} \left(\hat{\Sigma}_{\mu}(\mathsf{A}) + \hat{\Sigma}_{\mu}(\mathsf{B})\right)^{-1} \left(\hat{\mu}(\mathsf{A}) - \hat{\mu}(\mathsf{B})\right)$$

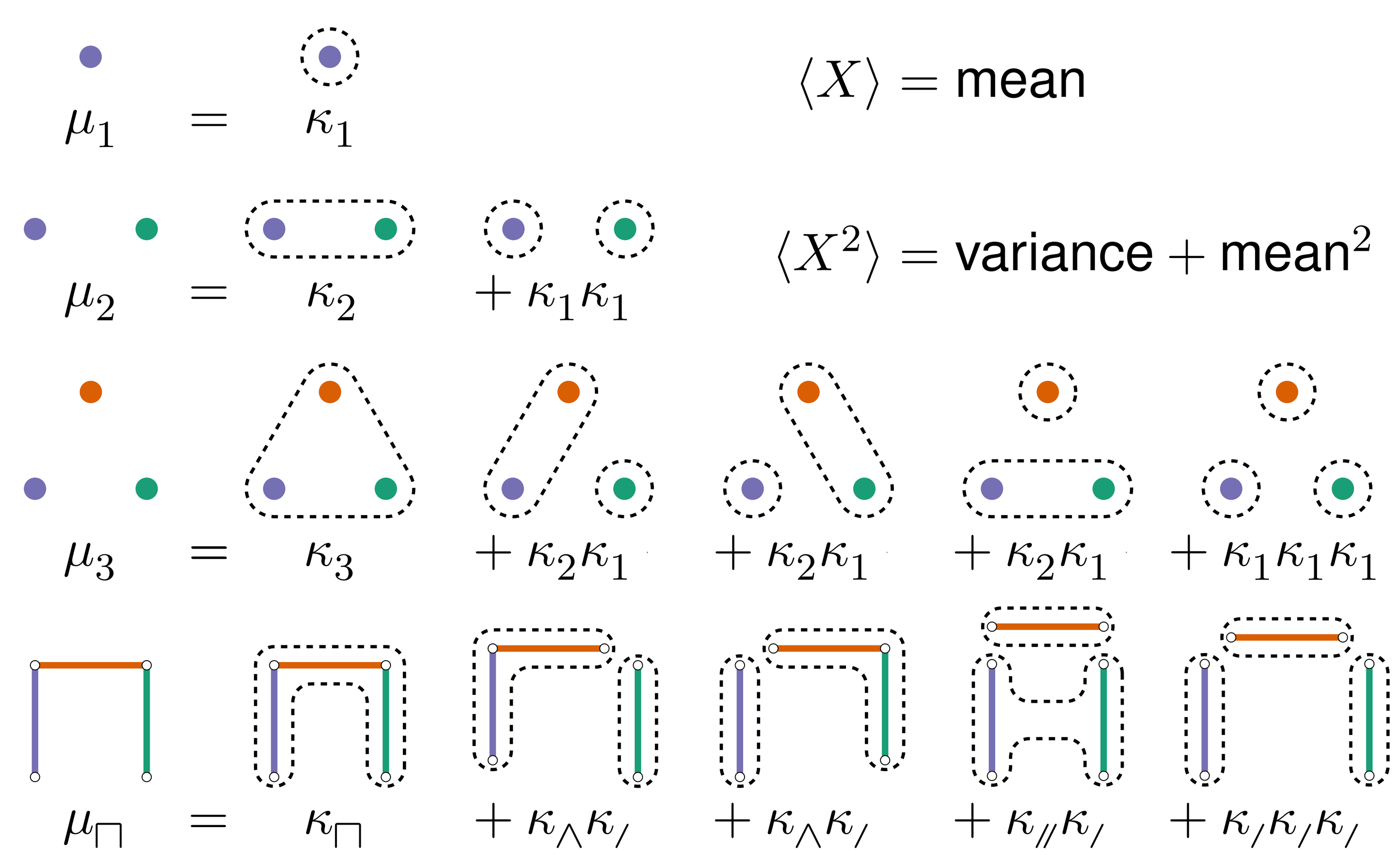
When estimating the covariance...

$$\mathrm{Cov}(\hat{\mu}_g, \hat{\mu}_{g'}) = \underbrace{\langle \hat{\mu}_g \, \hat{\mu}_{g'} \rangle}_{\text{“hard”}} - \underbrace{\langle \hat{\mu}_g \rangle \langle \hat{\mu}_{g'} \rangle}_{\text{“easy”}}$$

...the "hard" part uses a combinatorial disjoint union rule

$$c_{\wedge} c_{/} = 4c_{\wedge} + 2c_{\triangle} + 2c_{\lambda} + 4c_{\sqcap} + c_{/\!\!/\wedge}$$

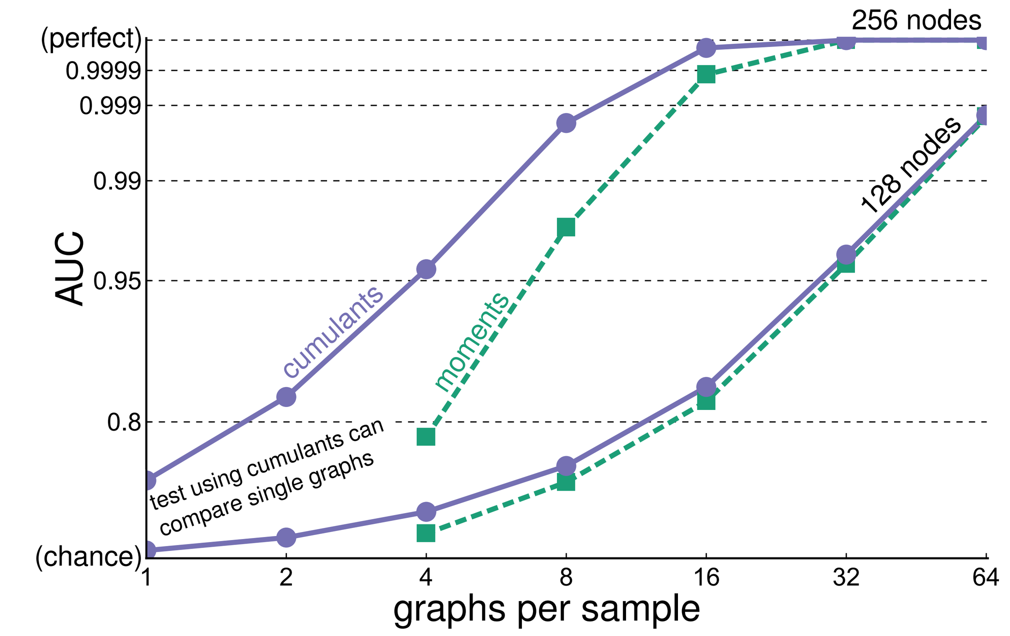
Combinatorial Construction of Cumulants

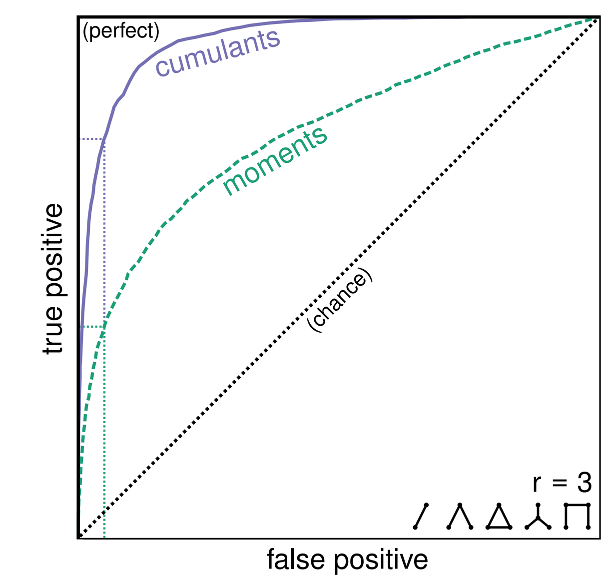


Graph Cumulants Clearly Conquer

Graph cumulants outperform subgraph densities in general...

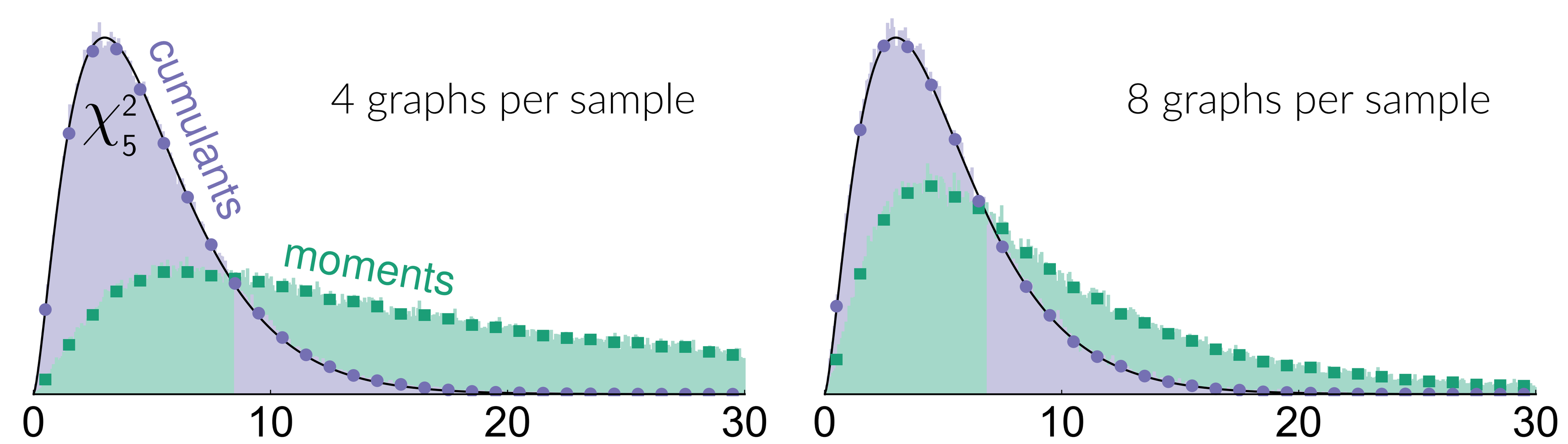
...and graph cumulants also work for single graph samples!





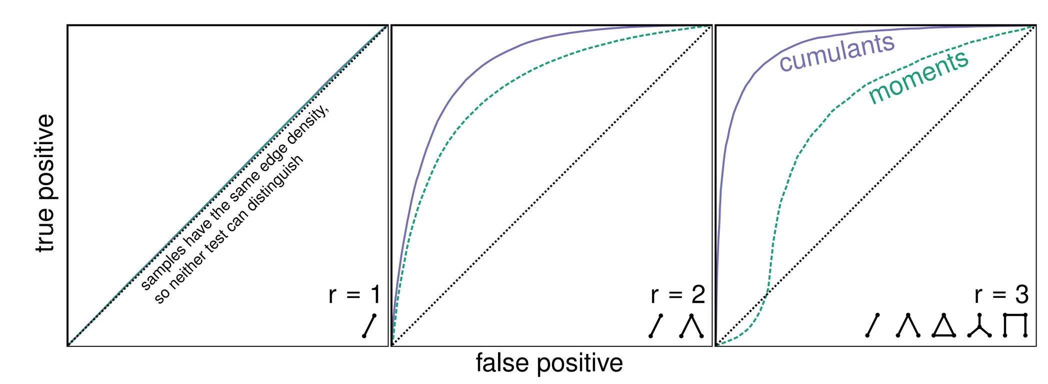
Why do graph cumulants perform better?

...because their **fluctuations** look more $\mathcal{N}$ormal!



"Z^2-score" for graphs sampled from the same distribution

Graph Cumulants in the (semi-)Wild!



Varying:
Number of subgraphs used by both tests

Comparing:
Genetic interaction networks of Arabidopsis and Mouse

Varying:
Number of graphs per sample used by both tests

Comparing:
Genetic interaction networks of Human and Rat