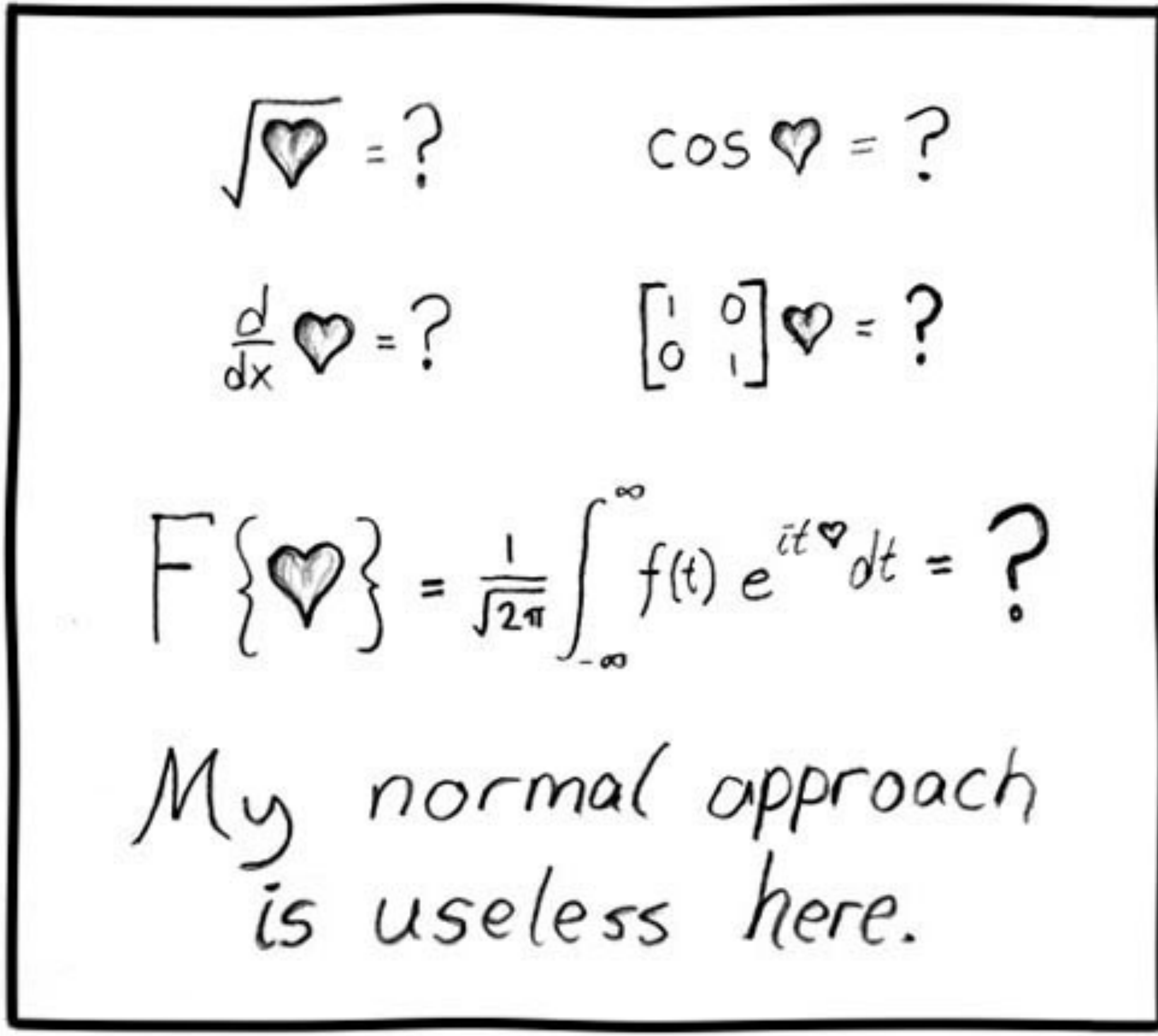
Non-commutative Harmonic Analysis in Machine Learning

Risi Kondor
Gatsby @ UCL / Caltech



http://www.xkcd.com/

$$e^{-i2\pi k x_1}\, e^{-i2\pi k x_2} = e^{-i2\pi k (x_1 + x_2)}$$

$$\widehat{f}(k) = \int e^{-i2\pi k x} f(x)\, dx$$

$$\rho(x_2)\rho(x_1) = \rho(x_2 x_1)$$

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\,\rho(x)$$

$\rho\colon G \to \mathbb{C}^{d \times d}$ is called a representation

Irreducible Representations of $\mathrm{SO}(3)$

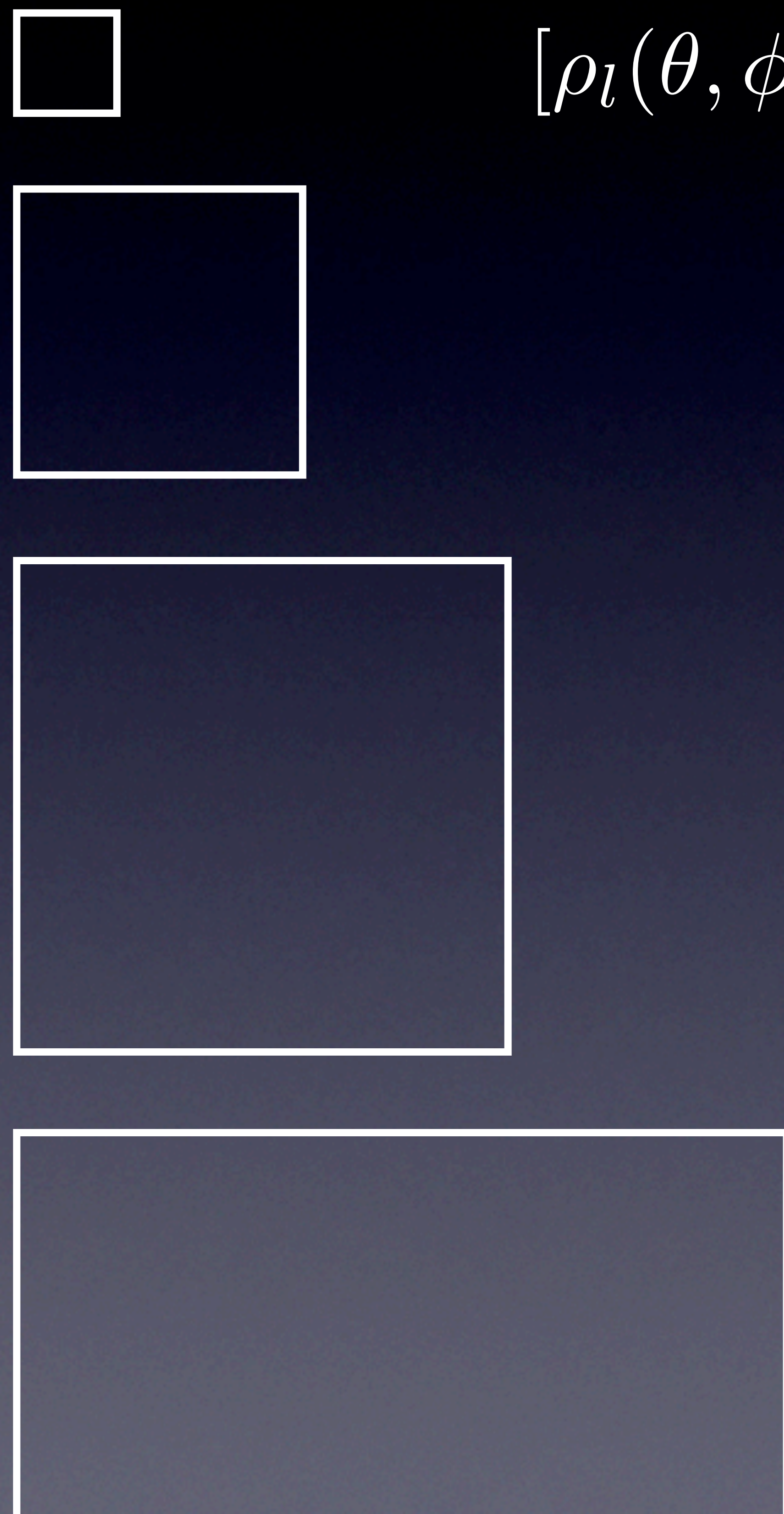
$$[\rho_l(\theta, \phi, \psi)]_{m,m'} = e^{-im'\psi} Y_l^m(\theta, \phi)$$

$\rho_0(R)$

$\rho_1(R)$

$\rho_2(R)$

$\rho_3(R)$



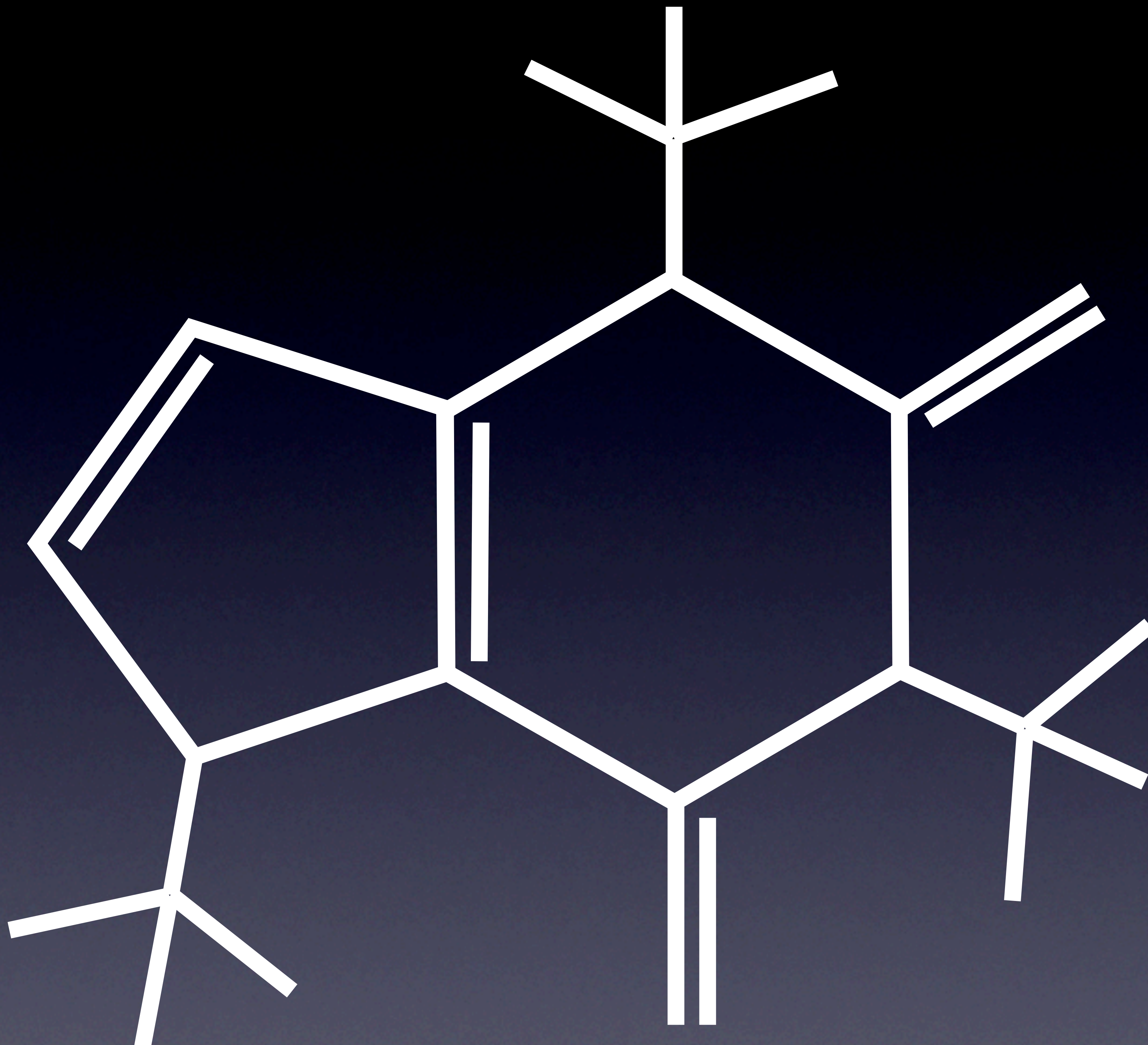
Group algebra: $\{\boldsymbol{e}_x\}_{x \in G}$ $\boldsymbol{e}_x \cdot \boldsymbol{e}_y := \boldsymbol{e}_{xy}$

$$\mathcal{F}: \; \mathbb{C}G \overset{isom.}{\longrightarrow} \bigoplus_{\rho \in \mathcal{R}} \mathrm{GL}(d_\rho)$$

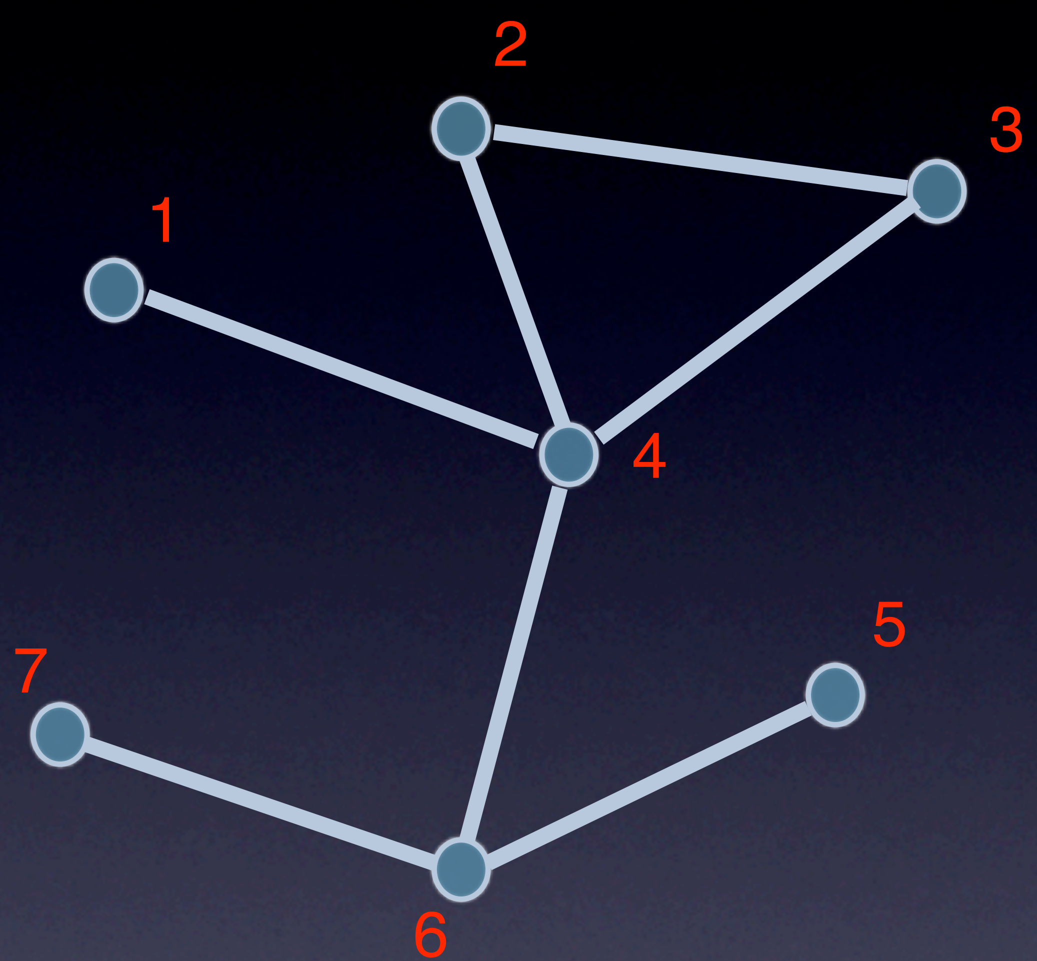
The skew spectrum and the graphlet spectrum of graphs

Karsten Borgwardt
Max Planck Inst., Tübingen

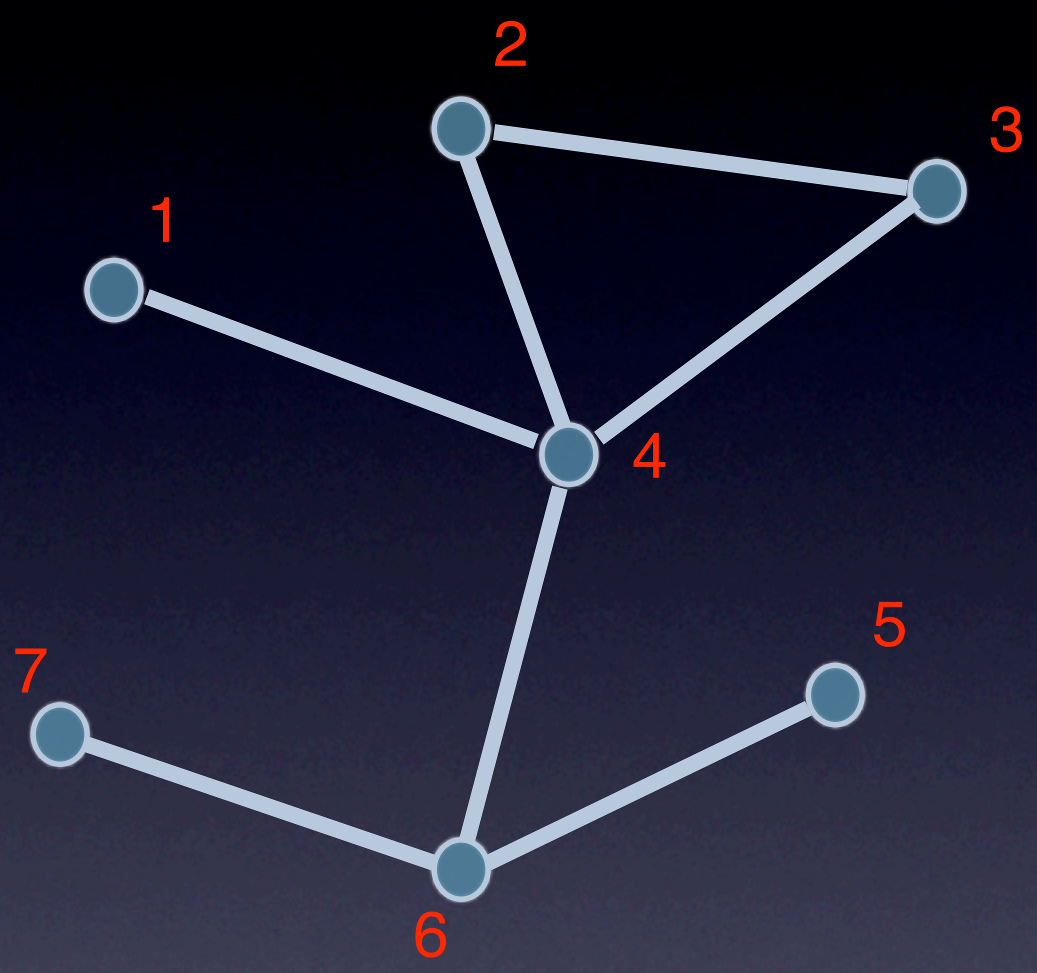
Nino Shervashidze
Max Planck Inst., Tübingen



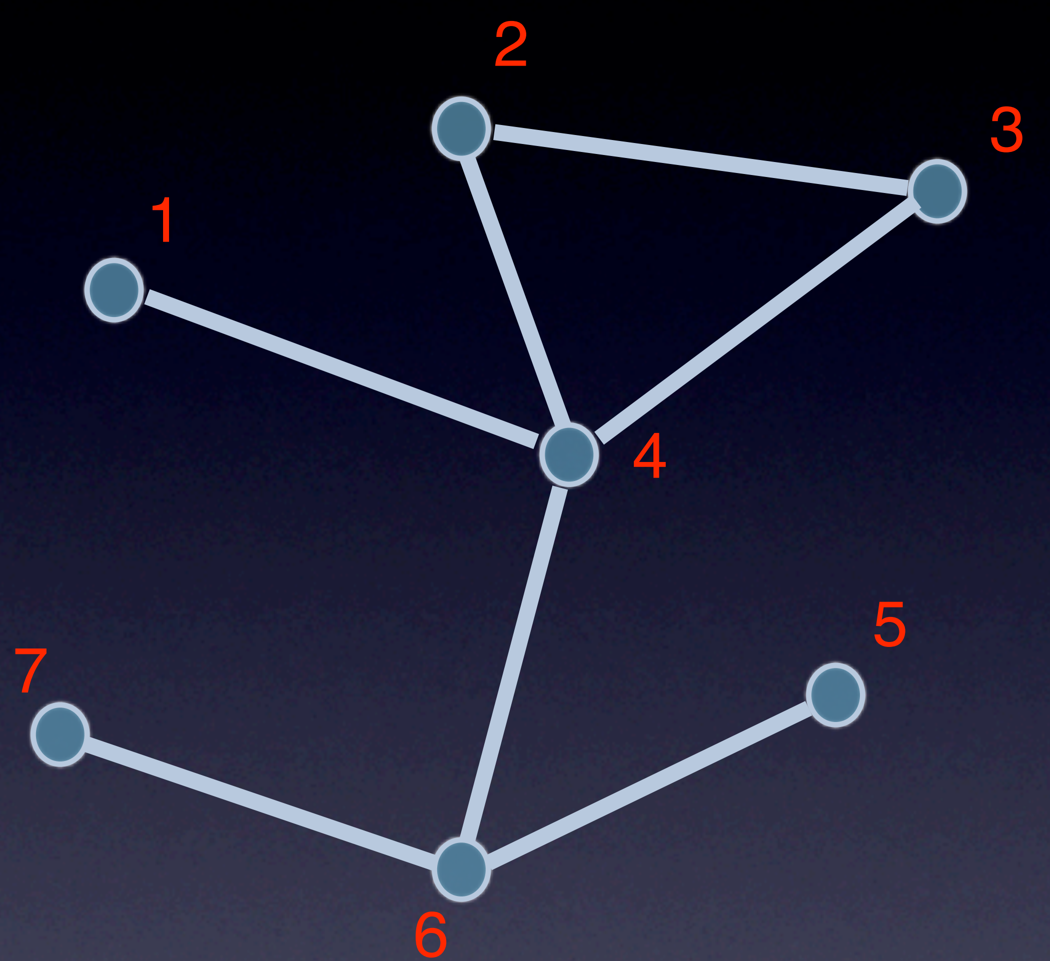
up to ~300 vertices



2
3
1
4
5
7
6



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$$A^{\pi}_{\pi(i),\pi(j)} = A_{i,j}$$

$q(A)$ is a graph invariant if $q(A^{\pi}) = q(A)$ for all π.

The graph function

$$f_A(\sigma) = [A]_{\sigma(n),\sigma(n-1)}$$

$$f_{A^\pi}(\sigma) = [A^\pi]_{\sigma(n),\sigma(n-1)} = A_{\pi^{-1}\sigma(n),\pi^{-1}\sigma(n-1)} = f_A(\pi^{-1}\sigma)$$

$f^\pi(\sigma) = f(\pi^{-1}\sigma)$ is called the left-translate of f

The Fourier spectrum of graphs

$$\widehat{f}_A(\rho) = \sum_{\sigma \in \mathbb{S}_n} f_A(\sigma)\, \rho(\sigma)$$

$$\widehat{f^\pi}(\rho) = \sum_{\sigma \in \mathbb{S}_n} f(\pi^{-1}\sigma)\, \rho(\sigma) = \sum_{\tau \in \mathbb{S}_n} f(\tau)\, \rho(\pi\tau) = \rho(\pi)\, \widehat{f}(\rho)$$

The power spectrum

$$\widehat{a}(\rho) = \widehat{f}(\rho)^\dagger \cdot \widehat{f}(\rho)$$

$$\widehat{a}^t(\rho) = (\rho(t)\widehat{f}(\rho))^\dagger \cdot (\rho(t)\widehat{f}(\rho)) = \widehat{f}(\rho)^\dagger \cdot \widehat{f}(\rho) = \widehat{a}(\rho)$$

The non-commutative bispectrum

$$\left(\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2)\right)^{\top} C_{\rho_1,\rho_2} \bigoplus_{\rho} \widehat{f}(\rho)$$

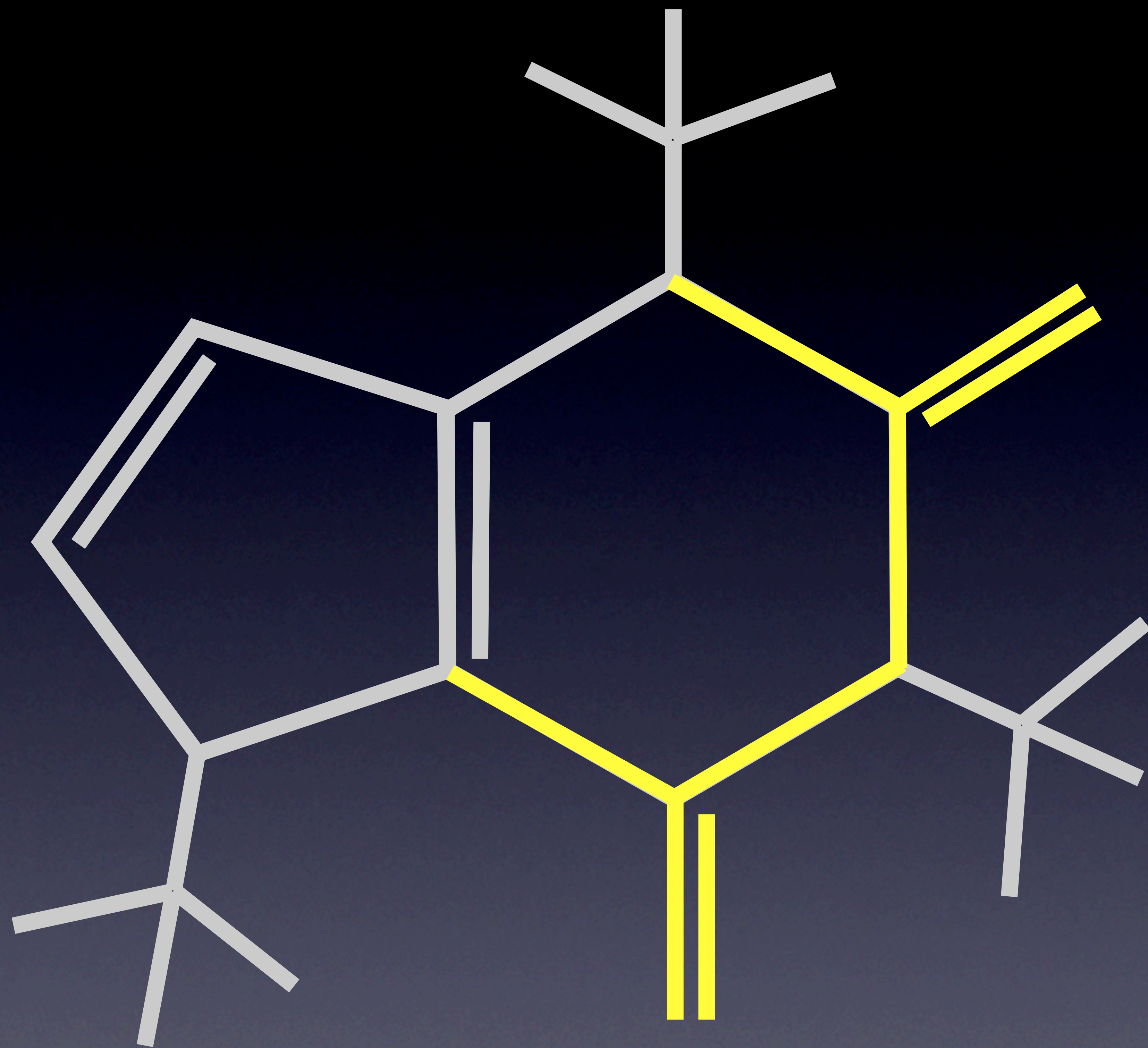
[Kakarala '92]

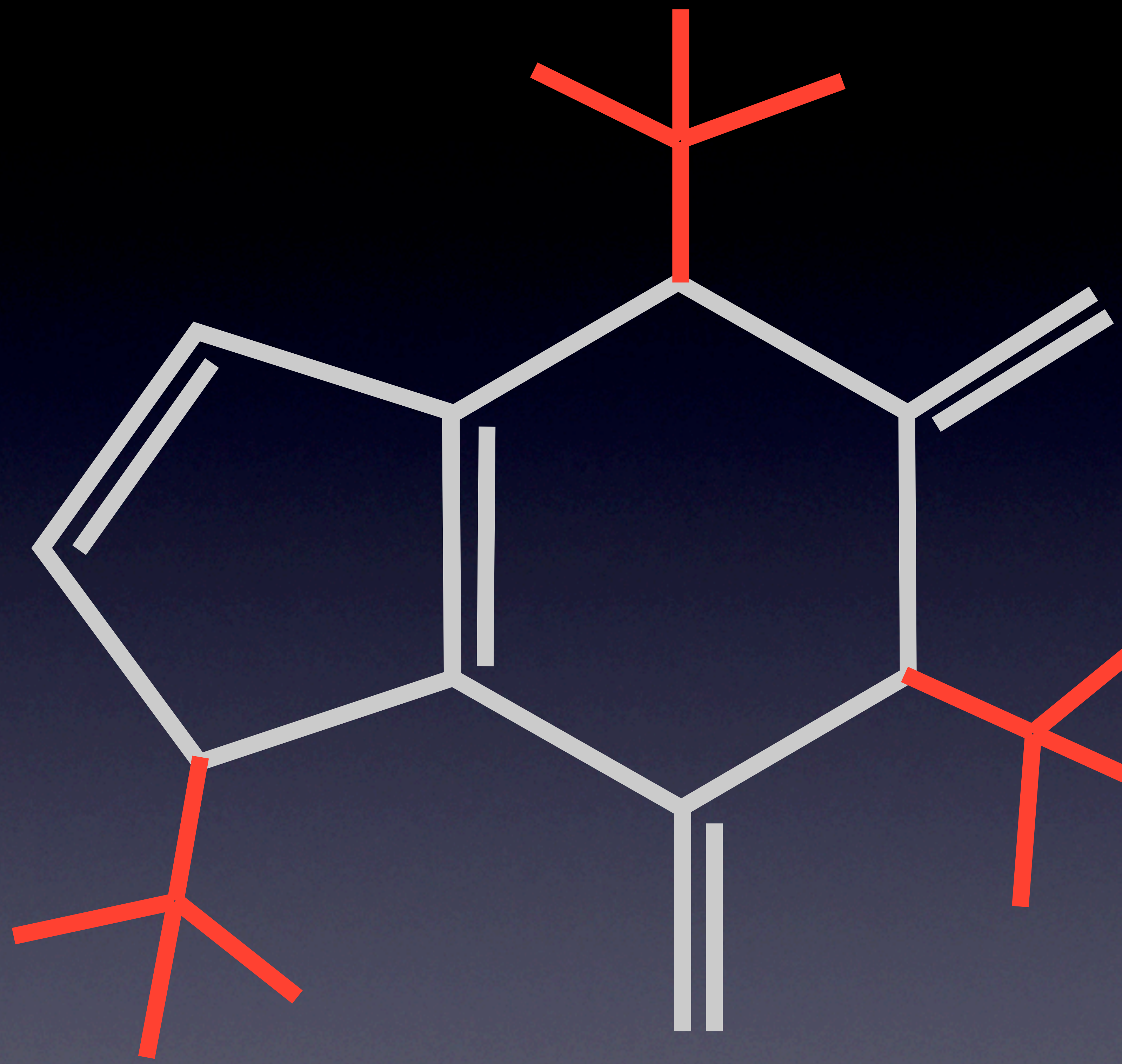
The skew spectrum

$$\widehat{q}_\nu(\rho) = \widehat{r}_\nu(\rho)^\top \cdot \widehat{f}(\rho)$$

$$r_\nu(\sigma) = f(\sigma\pi)f(\sigma)$$

[Kondor, 2007]





The graphlet spectrum

$f_g(\sigma) = 1$ if g occurs as a subgraph of
$\mathcal{G}$ at $(\sigma(n), \sigma(n-1), \ldots, \sigma(n-k+1))$

$$\widehat{f}_{g_i}(\lambda)^\dagger \cdot \widehat{f}_{g_j}(\lambda) \qquad i, j = 1, 2, \ldots m, \quad \lambda \vdash n$$

dirty details

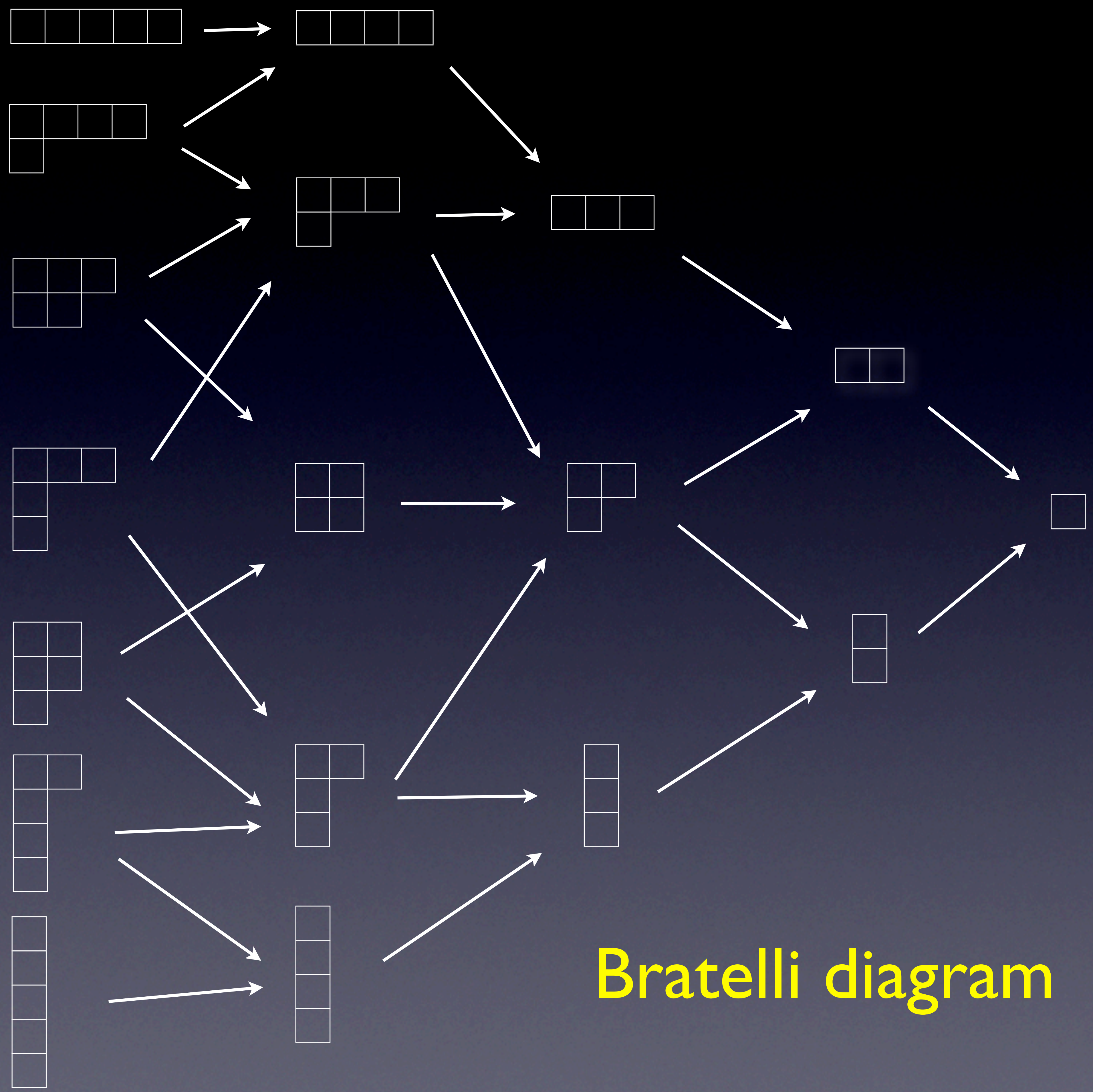
$$\sigma = (2, 1, 3, 4, 5) \in \mathbb{S}_5$$

$$\rho_1(\sigma) = \begin{pmatrix} 1 \end{pmatrix}$$

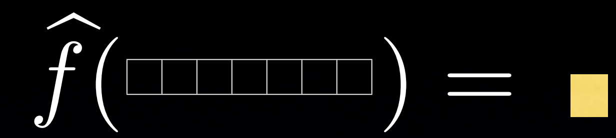
$$\rho_2(\sigma) = \begin{pmatrix} -0.5 & -0.289 & -0.204 & -0.791 \\ 0.866 & -0.167 & -0.118 & -0.456 \\ 0 & 0.943 & -0.0833 & -0.323 \\ 0 & 0 & 0.968 & -0.25 \end{pmatrix}$$

$$\rho_3(\sigma) = \begin{pmatrix} 0.25 & -0.433 & 0.433 & -0.75 & 0 \\ -0.433 & -0.25 & -0.75 & -0.433 & 0 \\ -0.433 & -0.25 & 0.25 & 0.144 & -0.816 \\ 0.75 & -0.144 & -0.433 & 0.0833 & -0.471 \\ 0 & 0.816 & 0 & -0.471 & -0.333 \end{pmatrix}$$

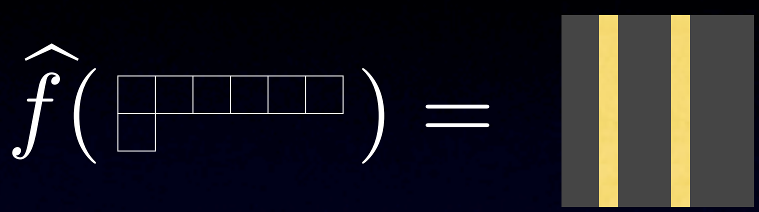
$$\rho_4(\sigma) = \begin{pmatrix} 0.333 & 0.236 & 0 & 0.913 & 0 & 0 \\ -0.471 & 0.0417 & 0.217 & 0.161 & 0.839 & 0 \\ 0.816 & -0.0722 & 0.125 & -0.28 & 0.484 & 0 \\ 0 & -0.484 & -0.28 & 0.125 & 0.0722 & 0.816 \\ 0 & 0.839 & -0.161 & -0.217 & 0.0417 & 0.471 \\ 0 & 0 & 0.913 & 0 & -0.236 & 0.333 \end{pmatrix}$$



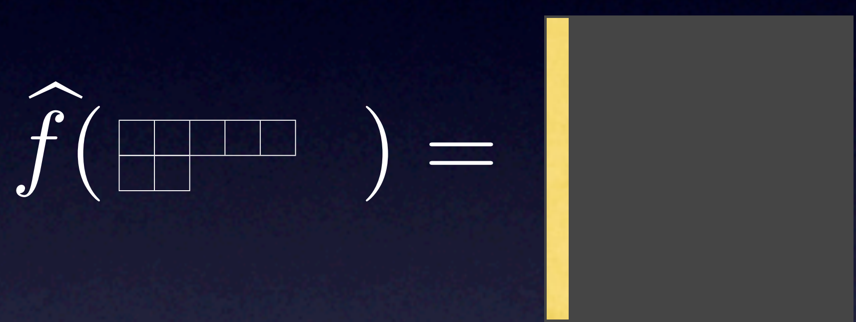
Bratelli diagram



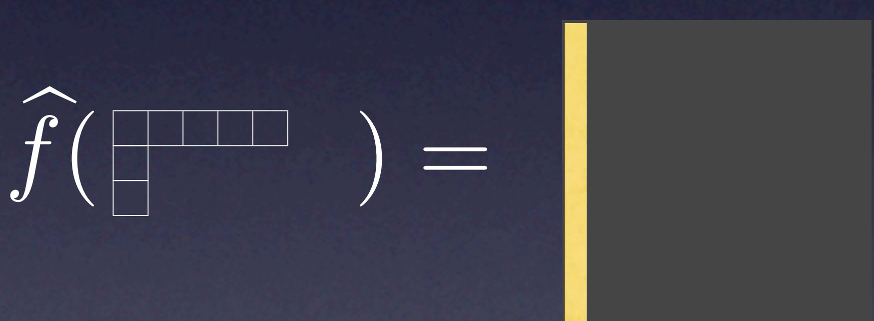
$d = 1$



$d = n - 1$



$d = n(n-3)/2$



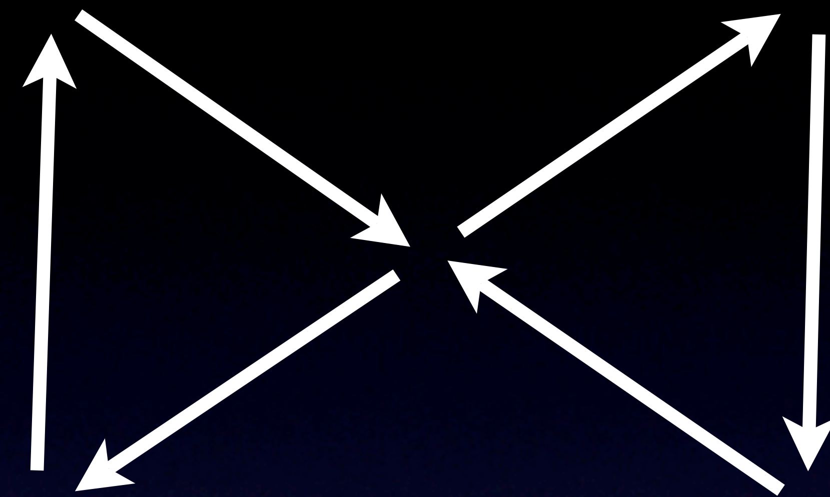
$d = (n-1)(n-2)/2$

$$\widehat{f}(\rho_1) = \boxed{(6)}$$

$$\widehat{f}(\rho_2) = \begin{pmatrix} -0.25 & -0.323 & 0 & 0 \\ -0.323 & -0.417 & 0 & 0 \\ 0.913 & 1.18 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\widehat{f}(\rho_3) = \begin{pmatrix} 1.33 & 0 & 0 & 0 & 0 \\ 0.471 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.816 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\widehat{f}(\rho_4) = \begin{pmatrix} -1.67 & 0 & 0 & 0 & 0 & 0 \\ 1.18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -0.913 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2.24 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



	size	time
Power spectrum	7	$O(n^3)$
Skew spectrum	87	$O(n^6)$
Reduced skew spectrum	49	$O(n^3)$
Graphlet spectrum	$\binom{m}{2} \sum_{s=0}^{k} \binom{k}{s} s!$	$O(n^{2+k})$

SnOB, FFT for the symmetric group

http://www1.cs.columbia.edu/~risi/SnOB/

$S_n$ob

A C++ library for fast Fourier transforms on the symmetric group.

author: Risi Kondor, Columbia University (risi@cs.columbia.edu)

Development version as of August 23, 2006 (unstable!):

Documentation: [ps][pdf]
C++ source code: [directory]
BiBTeX entry: [bib]
Entire package: [tar.gz]

References:

1. Michael Clausen: **Fast generalized Fourier transforms.** Theoretical Computer Science **67(1)**: 55-63, 1989.
2. David K. Maslen and Daniel N. Rockmore: **Generalized FFTs --- a survey of some recent results.** Proceedings of the DIMACS Workshop on Groups and Computation, 1997. [ps]
3. K.-L. Kueh, T. Olson, D. Rockmore and K.-S. Tan: **Nonlinear approximation theory on finite**

http://www.gatsby.ucl.ac.uk/~risi/SnOB

experimental results

	MUTAG	ENZYME	NCI1	NCI109
Number of instances/classes	600/6	188/2	4110/2	4127/2
Max. number of nodes	28	126	111	111
Reduced skew spectrum	**88.61** (0.21)	25.83 (0.34)	**62.72** (0.05)	**62.62** (0.03)
Random walk kernel	71.89 (0.66)	14.97 (0.28)	51.30 (0.23)	53.11 (0.11)
Shortest path kernel	81.28 (0.45)	**27.53** (0.29)	61.66 (0.10)	62.35 (0.13)

	MUTAG	ENZYMES	NCI1	NCI109
Number of instances/classes	188/2	600/6	4110/2	4127/2
Max. number of nodes	28	126	111	111
Graphlet spectrum	88.11 (0.46)	**35.42** (0.58)	**65.0** (0.09)	**65.31** (0.08)
Reduced skew spectrum	**88.61** (0.21)	25.83 (0.34)	62.72 (0.05)	62.62(0.03)
Graphlet count kernel	81.7 (0.67)	23.94 (0.4)	54.34 (0.04)	52.39 (0.09)

Multi-object tracking with representations of the symmetric group [K., Howard, Jebara, AISTATS 2007]

Efficient inference for distributions on permutations [Huang, Guestrin, Guibas, NIPS 2007]

The skew spectrum of graphs [K., Borgwardt, ICML 2008]

Characteristic kernels on groups and semigroups [Fukumizu, Sriperumbudur, Gretton, Scholkopf, NIPS 2008]

Inferring rankings under constrained sensing [Jagabathula, Shah, NIPS 2008]

The graphlet spectrum of graphs [K., Shervashidze, Borgwardt, ICML 2009]

...