



ALGEBRA IN DATA ANALYSIS

polynomial systems for statistics and biology

AIDA MARAJ



WHAT IS A POLYNOMIAL SYSTEM?

polynomials: expressions with numbers and variables, that involve only addition, subtraction and multiplication

$$x_1 + x_2 + x_3 - 1 \text{ (linear)}$$

$$56x_1x_2 \text{ (monomial)}$$

$$56x_1x_2 - 8x_3^2 \text{ (binomial)}$$

$$2x_1x_2^8 + 3x_1^2 + 6x_2^3x_3^4 - 7 \text{ (the others)}$$

not polynomials: 2^x , $\sin x$, $\frac{1}{y}$, ...



$$\begin{matrix} (p^2, & 2p(1-p), & (1-p)^2) \\ x_{20} & x_{11} & x_{02} \end{matrix}$$

$$4x_{20}x_{02} - x_{11}^2 \text{ (Jane's coin flip equation)}$$

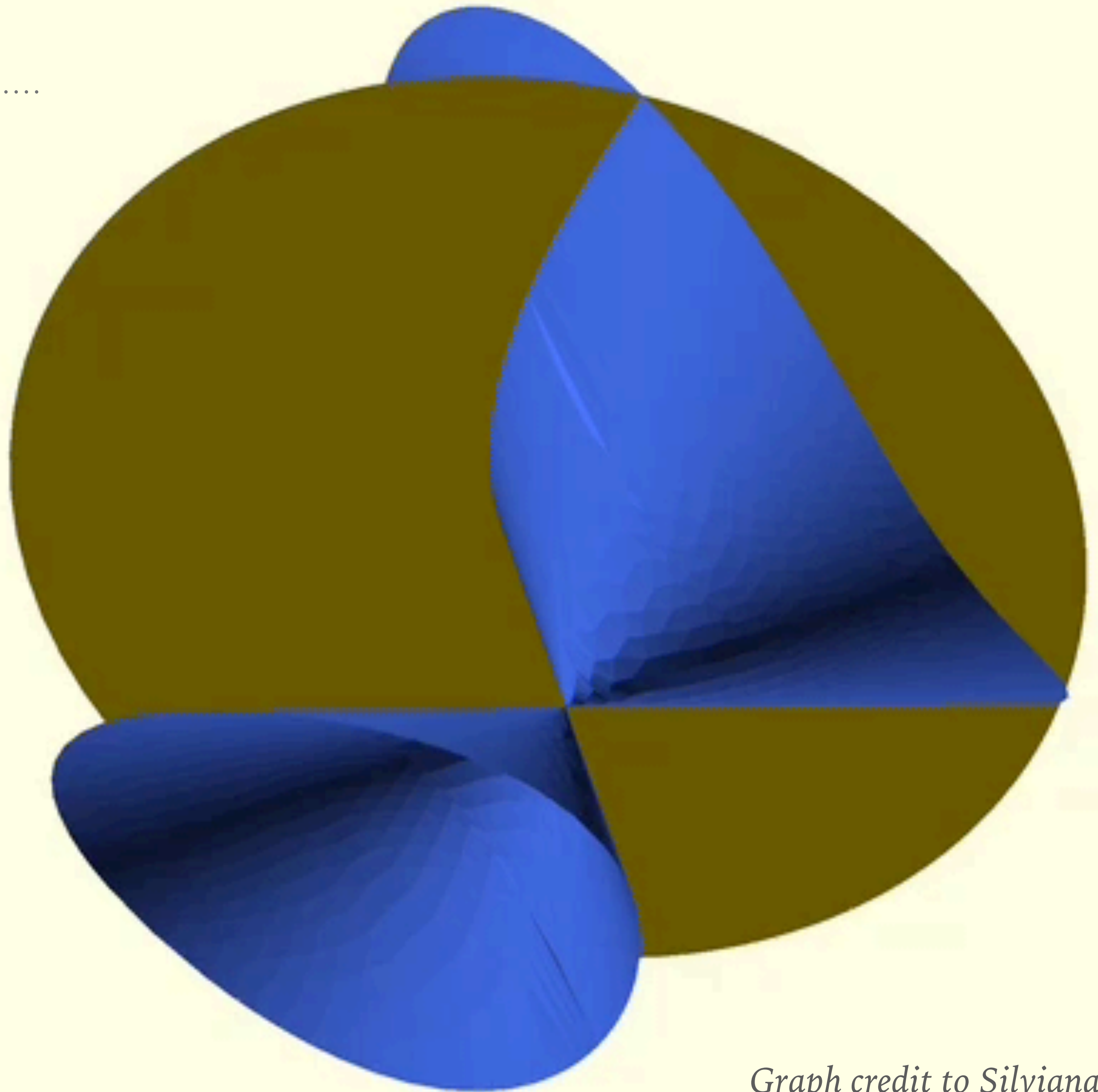
ZEROS OF A POLYNOMIAL SYSTEM

$$4x_1x_3 - x_2^2 = 0 \quad (1,2,1)$$

$$x_1 + x_2 + x_3 - 1 = 0 \quad (1,0,0)$$



$$(p^2, 2p(1-p), (1-p)^2)$$



Graph credit to Silvana Amethyst

DATA ANALYSIS VIA POLYNOMIALS

Take Aspirin Reg.	Age	Varicella	Previous Infection		
			Influenza	Gastroenteritis	
Yes	Child	19	11	2	
	Teen	9	10	1	
No	Child	400	100	74	
	Teen	304	88	50	

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

DATA ANALYSIS VIA POLYNOMIALS

Take Aspirin Reg.	Age	Previous Infection		
		Varicella	Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

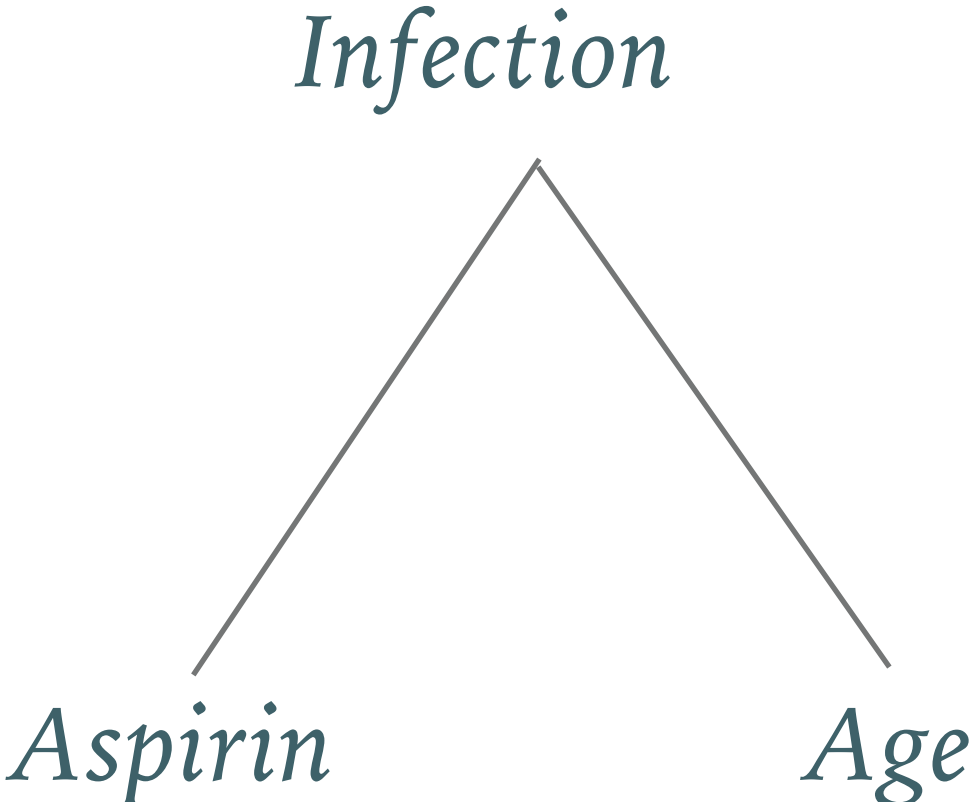
DATA ANALYSIS VIA POLYNOMIALS

Take Aspirin Reg.	Age	Varicella	Previous Infection	
			Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye’s syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

$$P(Inf = V, Asp = Y, Age = C) = P(Inf = V, Asp = Y) * P(Inf = V, Age = C)$$



DATA ANALYSIS VIA POLYNOMIALS

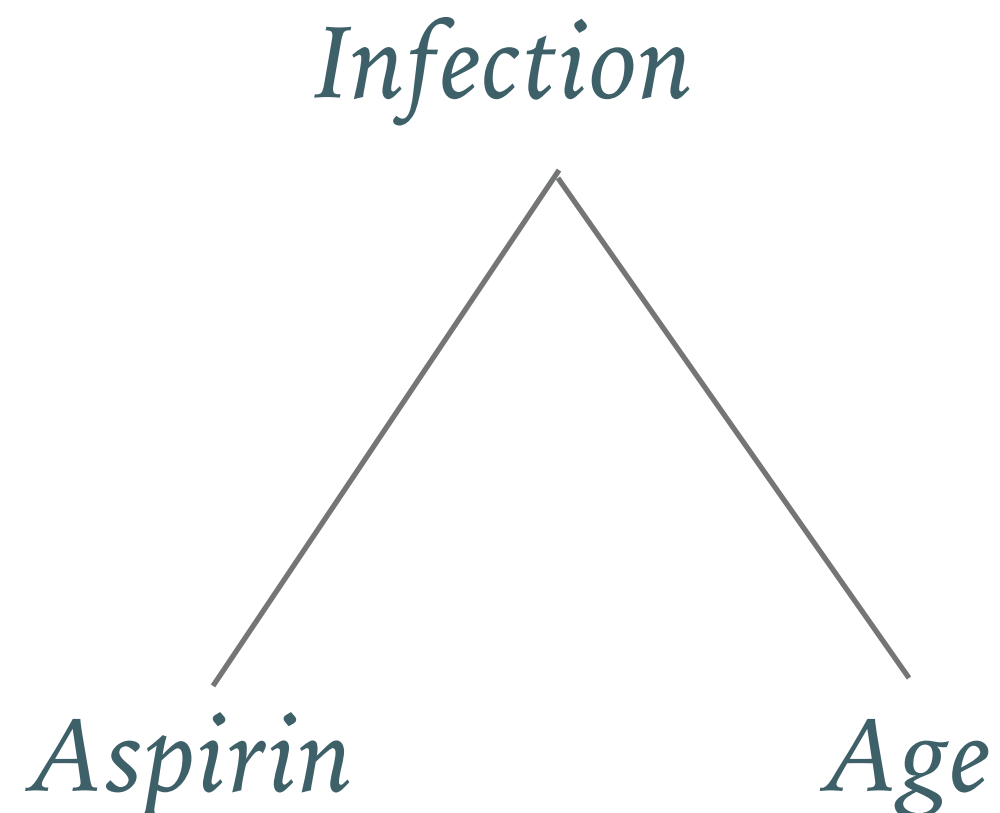
Previous Infection

Take Aspirin Reg.	Age	Varicella	Influenza	Gastroenteritis
Yes	Child	19/1070	11/1070	2/1070
	Teen	9/1070	10/1070	1/1070
No	Child	400/1070	100/1070	74/1070
	Teen	304/1070	88/1070	50/1070

*Belay, Bresee, Holman, Khan, Shahriari, and Schonberger. Reye's syndrome in the united states from 1981 through 1997. New England Journal of Medicine, 340(18):1377–1382, 1999.

Is there any relation between type of infection, taking aspirin regularly, and age?

$$P(\text{Inf} = V, \text{Asp} = Y, \text{Age} = C) = P(\text{Inf} = V, \text{Asp} = Y) * P(\text{Inf} = V, \text{Age} = C)$$

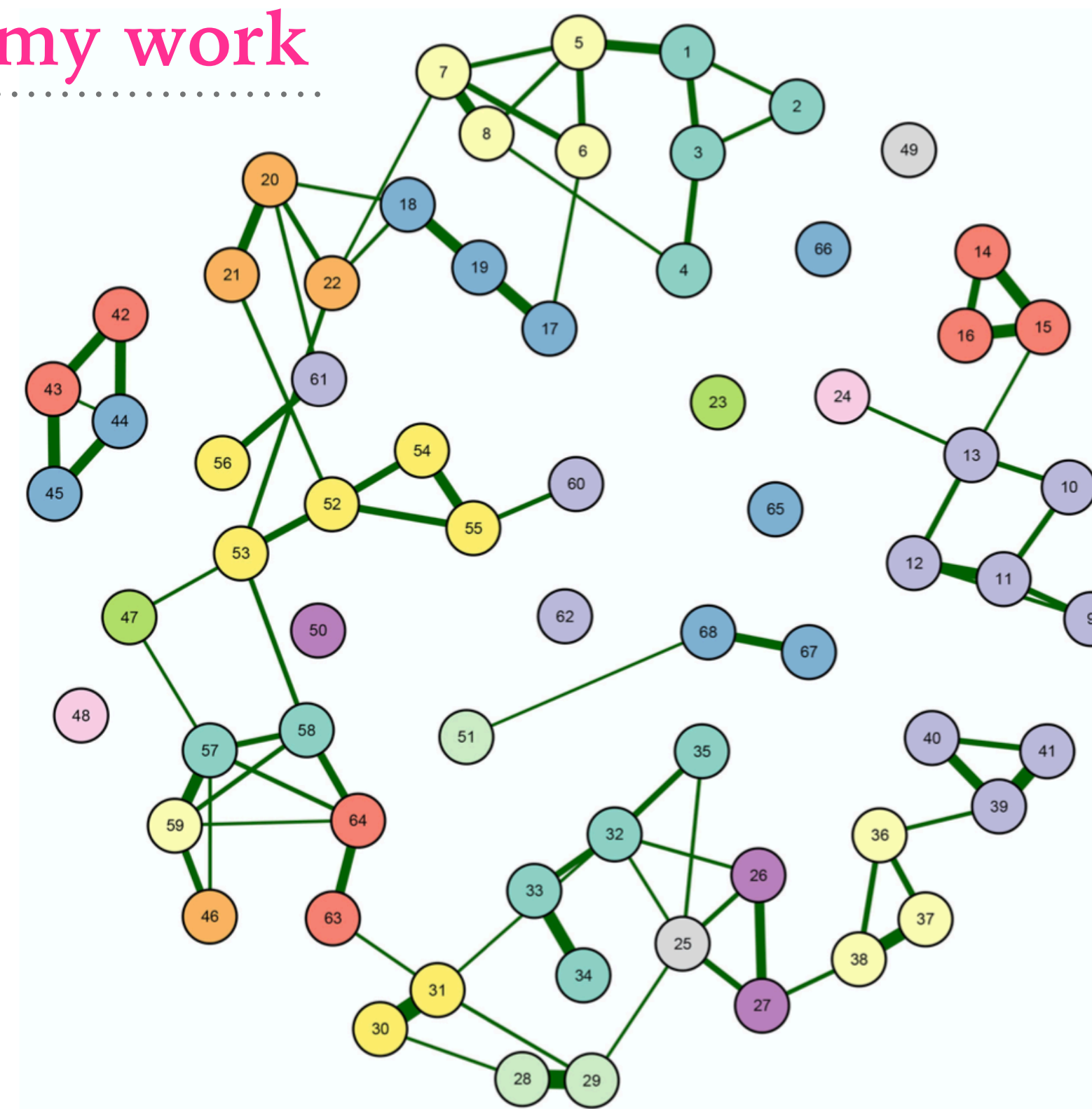
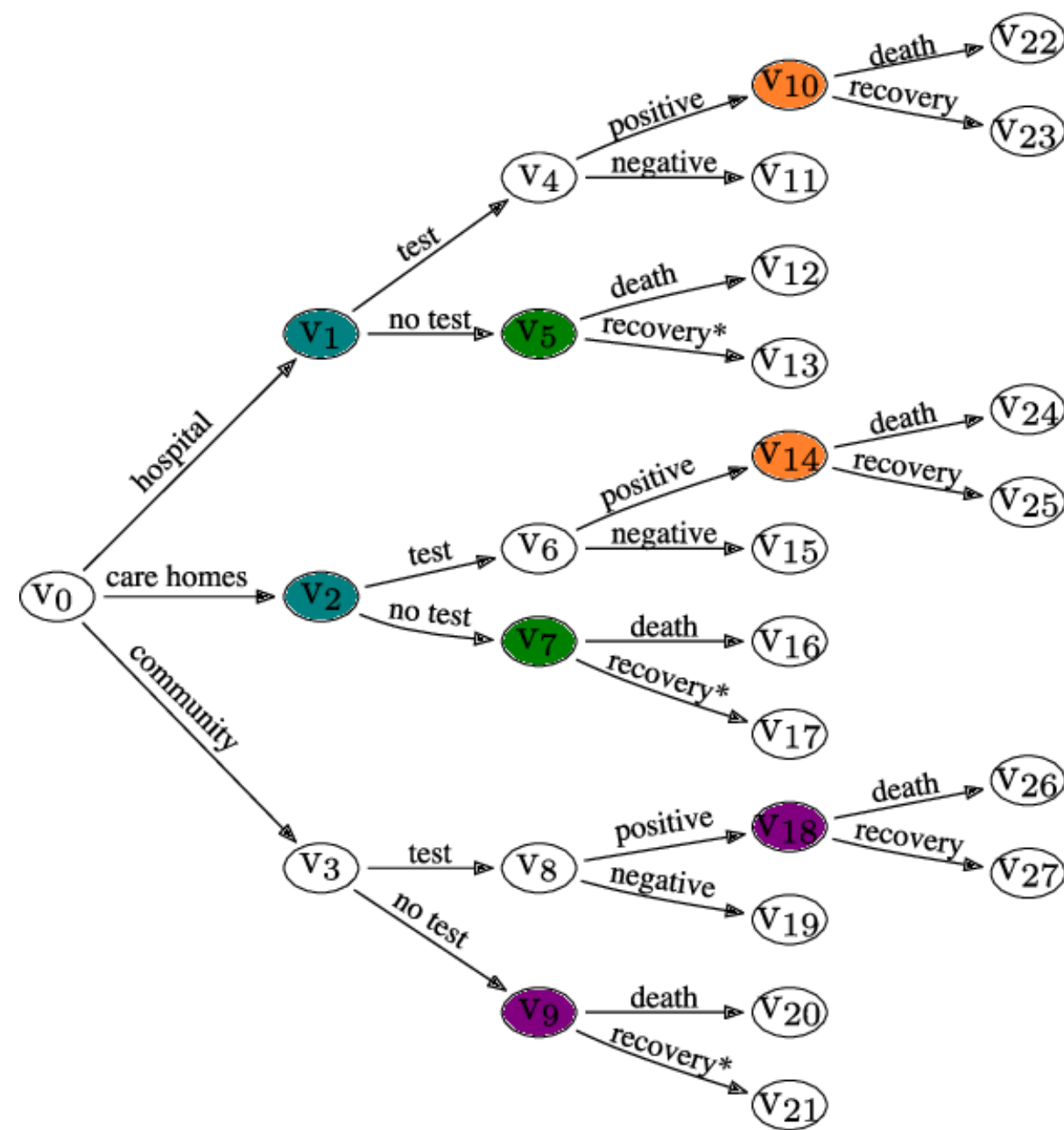


$$x_{111}x_{122} - x_{121}x_{112} = 0$$

$$x_{211}x_{222} - x_{221}x_{212} = 0$$

$$x_{311}x_{322} - x_{321}x_{312} = 0$$

Polynomial approaches to statistical models in my work



- Altruistic values [1–4]
- Biospheric values [5–8]
- Egoistic values [9–13]
- Hedonic values [14–16]
- Environmental self-identity [17–19]
- Personal importance of sustainable energy behaviour [20–22]
- Need to belong [23]
- Need to be unique [24]
- Neighbourhood entitativity [25]
- Neighbourhood homogeneity [26–27]
- Neighbourhood interaction [28–29]
- Interaction with neighbours [30–31]
- Neighbourhood identification [32–35]
- Environmental neighbourhood identity [36–38]
- Neighbourhood importance of sustainable energy behaviour [39–41]
- Group-based anger [42–43]
- Group-based distrust [44–45]
- Membership [46]
- Overall energy savings [47]
- Thermostat temperature (°C) [48]
- Shower time (min) [49]
- Energy-efficient appliances [50]
- Energy-saving measures [51]
- Household sustainable energy intentions [52–56]
- Communal sustainable energy intentions [57–58]
- Initiative involvement intentions [59]
- Other pro-environmental intentions [60–62]
- Other communal intentions [63–64]
- Demographical variables [65–68]

Polynomial approaches to discrete models

Taylor Brysiewicz and Aida Maraj. **Lawrence lifts, matroids, and maximum likelihood degrees.** 2023.

Jane Ivy Coons, Joseph Cummings, Benjamin Hollering, and Aida Maraj. **Generalized cut polytopes for binary hierarchical models.** *Algebraic Statistics*, 2023

Christiane Görden, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.

Aida Maraj. **Algebraic and geometric properties of hierarchical models.** *PhD dissertation—University of Kentucky*, 2020

Aida Maraj and Uwe Nagel. **Equivariant Hilbert series for hierarchical models.** *Algebraic Statistics*, 12(1):21–42, 2021.

Polynomial approaches to multivariate Gaussian models

Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models.** *Le Matematiche*, 76(2):383 – 398, 2021.

Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance,** 2024

Jane Ivy Coons, Aida Maraj, Pratik Misra, and Miruna-Stefana Sorea. **Symmetrically colored Gaussian graphical models with toric vanishing ideals.** *SIAM Journal on Applied Algebra and Geometry*, 7(1):133–158, 2023.

Authors listed alphabetically

Polynomial approaches in theory of evolution dates back to 1893

III. Contributions to the Mathematical Theory of Evolution.

By KARL PEARSON, University College, London.

Communicated by Professor HENRICI, F.R.S.

Received October 18,—Read November 16, 1893.

[PLATES 1—5.]

CONTENTS.

	Page.
I.—On the Dissection of Asymmetrical Frequency-Curves. General Theory, §§ 1–8.	71–85
Example: Professor WELDON'S measurements of the "Forehead" of Crabs.	
§§ 9–10	85–90
II.—On the Dissection of Symmetrical Frequency-Curves. General Theory, §§ 11–12	
Application. Crabs "No. 4," §§ 13–15	90–100
III.—Investigation of an Asymmetrical Frequency-Curve representing Mr. H. THOMSON'S	
measurements of the Carapace of Prawns. §§ 16–18	100–106
Table I. First Six Powers of First Thirty Natural Numbers	106
Table II. Ordinates of Normal Frequency-Curve	107
Note added February 10, 1894	107–110

...to explain the asymmetry in data measured from a population of Naples' crabs, believing it was possible that two subpopulations of crabs were present in the sample — Gaussian mixture. He proposes method of moments.

$$\alpha_1 + \alpha_2 = 1$$

$$\alpha_1\mu_1 + \alpha_2\mu_2 = 0$$

$$\alpha_1(\mu_1^2 + \sigma_1^2) + \alpha_2(\mu_2^2 + \sigma_2^2) = m_2$$

$$\alpha_1(\mu_1^3 + 3\mu_1\sigma_1^2) + \alpha_2(\mu_2^3 + 3\mu_2\sigma_2^2) = m_3$$

$$\alpha_1(\mu_1^4 + 6\mu_1^2\sigma_1^2 + 3\sigma_1^4) + \alpha_2(\mu_2^4 + 6\mu_2^2\sigma_2^2 + 3\sigma_2^4) = m_4$$

$$\alpha_1(\mu_1^5 + 10\mu_1^3\sigma_1^2 + 15\mu_1\sigma_1^4) + \alpha_2(\mu_2^5 + 10\mu_2^3\sigma_2^2 + 15\mu_2\sigma_2^4) = m_5.$$

THEORY OF POLYNOMIAL SYSTEMS = ALGEBRAIC GEOMETRY

algebra software: Julia, Macaulay2, Bertini, SageMath, Oscar

a trick: invertible linear change of variables

$$x_1x_2 - x_3x_4 - x_5x_4 + x_1x_6 = x_1(x_2 + x_6) - (x_3 + x_5)x_4 = p_1p_2 - p_3p_4$$

Relevant theory and applications in my past work

1. Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models.** *Le Matematiche*, 76(2):383 – 398, 2021.
2. Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance**, 2024
3. Christiane Görgen, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.
4. Aida Maraj and Arpan Pal. **Symmetry Lie algebras of varieties with applications to algebraic statistics**, 2023.
5. Thomas Kahle and Julian Vill. **Efficiently deciding if an ideal is toric after a linear coordinate change.** 2024.

THEORY OF POLYNOMIAL SYSTEMS = ALGEBRAIC GEOMETRY

algebra software: Julia, Macaulay2, Bertini, SageMath, Oscar

a trick: invertible linear change of variables

$$x_1x_2 - x_3x_4 - x_5x_4 + x_1x_6 = x_1(x_2 + x_6) - (x_3 + x_5)x_4 = p_1p_2 - p_3p_4$$

Relevant theory and applications in my past work

1. Tobias Boege, Jane Ivy Coons, Christopher Eur, Aida Maraj, and Frank Röttger. **Reciprocal maximum likelihood degrees of Brownian motion tree models.** *Le Matematiche*, 76(2):383 – 398, 2021.
2. Jane Ivy Coons, Shelby Cox, Aida Maraj, and Ikenna Nometa. **Maximum Likelihood Degrees of Brownian Motion Tree Models: Star Trees and Root Invariance**, 2024
3. Christiane Görgen, Aida Maraj, and Lisa Nicklasson. **Staged tree models with toric structure.** *Journal of Symbolic Computation*, 113:242–268, 2022.
4. Aida Maraj and Arpan Pal. **Symmetry Lie algebras of varieties with applications to algebraic statistics**, 2023.
5. Thomas Kahle and Julian Vill. **Efficiently deciding if an ideal is toric after a linear coordinate change.** 2024.

The Annals of Statistics
1993, Vol. 21, No. 1, 355–377

INVARIANTS OF SOME PROBABILITY MODELS USED IN PHYLOGENETIC INFERENCE

BY STEVEN N. EVANS¹ AND T. P. SPEED²

University of California, Berkeley

The so-called method of invariants is a technique in the field of molecular evolution for inferring phylogenetic relations among a number of species on the basis of nucleotide sequence data. An invariant is a polynomial function of the probability distribution defined by a stochastic model

THE ALGEBRA IN DATA ANALYSIS GROUP

at MPI CBG and CSBD will:

- investigate statistical and biological questions through the lens of algebra, geometry and combinatorics, and
- develop new math motivated by this perspective

THE ALGEBRA IN DATA ANALYSIS GROUP

at MPI CBG and CSBD will:

- investigate statistical and biological questions through the lens of algebra, geometry and combinatorics, and
- develop new math motivated by this perspective

THANK YOU!