MSRI
Mathematical Sciences Research Institute

17 Gauss Way Berkeley, CA 94720-5070 p: 510.642.0143 f: 510.642.8609 www.msri.org

NOTETAKER CHECKLIST FORM

(Complete one for each talk.)

Name: Elizabeth Gross Email/Phone: egross7@uic.edu

Speaker's Name: Sonja Petrović

Talk Title: Toric geometry of hypergraphs

Date: 12 / 4 / 12 Time: 9 : 00 (am) / pm (circle one)

List 6-12 key words for the talk: hypergraphs, toric ideals, edge subring, Graver basis, social networks, Markov basis

Please summarize the lecture in 5 or fewer sentances: Introduces the presentation ideals of edge subrings of hypergraphs. Describes the combinatorics of primitive & indispensable binomals in the toric ideal of a hypergraph. As an application to algebraic statistics, introduces a statistical model for social networks.

CHECK LIST

(This is **NOT** optional, we will **not pay** for **incomplete** forms)

- ☐ Introduce yourself to the speaker prior to the talk. Tell them that you will be the note taker, and that you will need to make copies of their notes and materials, if any.
- ☐ Obtain ALL presentation materials from speaker. This can be done before the talk is to begin or after the talk; please make arrangements with the speaker as to when you can do this. You may scan and send materials as a .pdf to yourself using the scanner on the 3rd floor.
  - <u>Computer Presentations</u>: Obtain a copy of their presentation
  - <u>Overhead</u>: Obtain a copy or use the originals and scan them
  - <u>Blackboard</u>: Take blackboard notes in black or blue **PEN**. We will **NOT** accept notes in pencil or in colored ink other than black or blue.
  - <u>Handouts</u>: Obtain copies of and scan all handouts
- ☐ For each talk, all materials must be saved in a single .pdf and named according to the naming convention on the "Materials Received" check list. To do this, compile all materials for a specific talk into one stack <u>with this completed sheet on top</u> and insert face up into the tray on the top of the scanner. Proceed to scan and email the file to yourself. Do this for the materials from each talk.
- ☐ When you have emailed all files to yourself, please save and re-name each file according to the naming convention listed below the talk title on the "Materials Received" check list. (*YYYY.MM.DD.TIME.SpeakerLastName*)
- ☐ Email the re-named files to notes@msri.org with the workshop name and your name in the subject line.

Toric ideals of hypergraphs

- and applications to algebraic statistics -

Sonja Petrović

Statistics Department
Penn State University

Joint work with Elizabeth Gross and Despina Stasi

Combinatorial Commutative Algebra and Applications
3-7 December 2012
MSRI

arXiv:1206.1904, arXiv:1206.2512

Overview

Monomial algebras associated to graphs

Fundamental algebraic properties $\leftrightarrow$ combinatorics

(Simis-Vasconcelos-Villarreal '94, ...)

Presentation ideals of edge subrings of graphs

Parametrized by monomial maps whose images are edges of a graph.

(Villarreal, Ohsugi-Hibi, Reyes-Tatakis-Thoma)

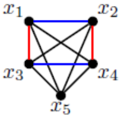
- Toric ideals virtually unexplored beyond the graph case
- Of interest as blow-up algebras (the edge subring of a uniform hypergraph is isomorphic to the special fiber ring of its edge ideal)

This talk is an overview of algebraic results with impact in algebraic statistics, motivated by their applied relevance.

Edge subring of a hypergraph

The edge subring of the hypergraph H, denoted by $k[H]$, is the monomial subring generated by the edges of H:

$$k[H] := k[x^{e_i} : e_i \in E(H)].$$



Let t_{e_i} be a variable representing the edge e_i. ($t_{13} \leftrightarrow$ edge $x_1 x_3$.)
The toric ideal I_H of the edge subring of the hypergraph H is the kernel of the monomial map ϕ_H:

$$\phi_H : k[t_{e_i}] \to k[H]$$
$$t_{e_i} \mapsto x^{e_i}.$$

Example

I_{K_5} is generated by quadrics of the form $t_{13}t_{24} - t_{12}t_{34}$, arising from cycles in K_5.

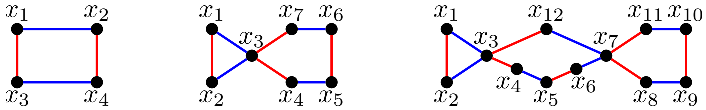
Monomial walks on graphs

Theorem (Ohsugi-Hibi '99-'00, Villarreal '01)

The toric ideal I_G is generated by binomials arising from (primitive) even closed walks on G.

Theorem (Ohsugi-Hibi '99-'00, Villarreal '01, Reyes-Tatakis-Thoma '12)

Primitive even closed walks on G are:
(i) even cycles,
(ii) two odd cycles sharing a vertex, or
(iii) two odd cycles such that there are two walks connecting a vertex in one cycle with a vertex in the other.



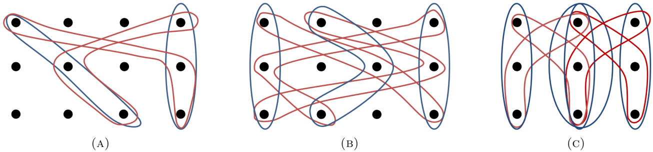
The three types of walks characterizing generators of I_G

Monomial walks on hypergraphs

An even collection $\mathcal{E}$ of edges in H is balanced with respect to a given bicoloring of H if

$$\deg_{blue}(v) = \deg_{red}(v) \qquad (\star)$$

for each vertex v covered by $\mathcal{E}$.

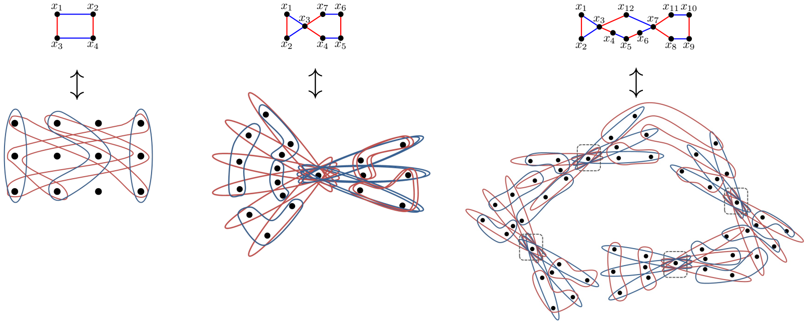


Three balanced sets on a 3-uniform hypergraph

Theorem (P.-Stasi '12)

Any balanced collection of edges constitutes a monomial walk.
In particular, the ideal I_H is generated by binomials arising from primitive monomial walks on H. They form the Graver basis.

Combinatorics of sparse bouquets



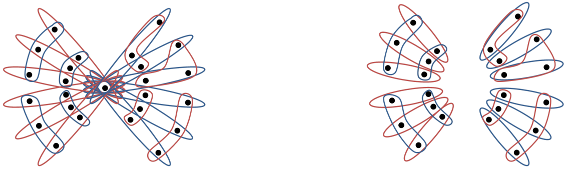
Theorem (P.-Stasi '12)

Let $H =$ bouquet of ℓ sunflowers with matching on vertices in $H - C$.
H supports a binomial in I_H if and only if there exists a partition $(G_{\mathcal{J}}, G_{\mathcal{K}})$ of the connected components of $H - C$ such that for each $1 \leq i \leq \ell$:

$$\sum_{j \in \mathcal{J}(S_i)} m_j |G_j| = \sum_{k \in \mathcal{K}(S_i)} m_k |G_k|, \quad \textit{where } m_j,\, m_k \textit{ are integers.}$$

Graver basis of I_H: detecting primitivity

- In ‘sparse’ cases, counting sunflower petals (with mutliplicities) detects balanced sets on H, and therefore, primitivity.



- For arbitrary binomials, there is no efficient way to detect primitivity.
- Problem: finding good sunflower decompositions.

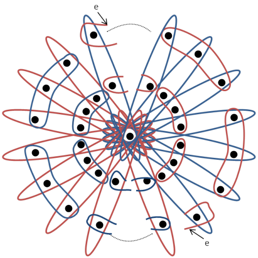
For *squarefree* binomials in I_H, detecting primitivity is dual to the *discrepancy* problem in hypergraphs.

Bouquet complexity

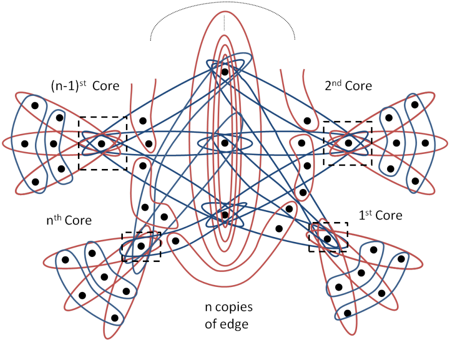
In the case of graphs:

- the largest degree of a vertex in the support of a primitive walk is 4
- the edges in the support of the walk can be traversed at most twice.

None of these restrictions extend to d-uniform hypergraphs ($d > 2$).



A primitive monomial hypergraph containing a vertex with arbitrarily large degree.



A primitive monomial hypergraph with arbitrarily many cores.

The motivating problem: fitting network models to data

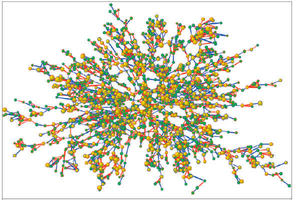


Figure 1. Largest Connected Subcomponent of the Social Network in the Framingham Heart Study in the Year 2000.

Does a proposed log-linear model fit this observed network?

- This can be answered by computing a Markov basis for the model, which is used for a random walk on a fiber.

- In algebraic statistics, we study statistical models whose parameter spaces correspond to real positive parts of algebraic varieties:

Theorem (Diaconis-Sturmfels '98)

Markov bases are generating sets of the toric ideal of the model.

Fact:

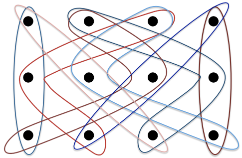
Toric ideals of many popular statistical models for networks are presentations ideals of edge subrings of hypergraphs.

Markov bases $\leftrightarrow$ balanced hypergraphs

- Markov basis for the model $\mathcal{M}$ is thus described by bicolored monomial hypergraphs over the model hypergraph $H_{\mathcal{M}}$:

If $\mathcal{E}$ supports a binomial over $H_{\mathcal{M}}$, then a Markov move on a fiber of the model corresponds to replacing the set of red edges in $\mathcal{E}$ by the set of blue edges in $\mathcal{E}$.

- Degree bounds for minimal generators $\rightarrow$ a bound for the *Markov complexity* (Markov width) of the model $\mathcal{M}$.
- Sometimes, the squarefree part of the Graver basis of I_H is required for the full Markov basis (Hara-Takemura '10).



$H_{\mathcal{M}}$ for the no 3-way interaction model

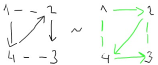
The story of 80,000 Markov moves

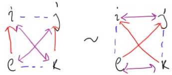
Theorem (*P.*-Rinaldo-Fienberg '10)

The toric ideal of the p_1 random graph model on n nodes is the multi-homogenous piece of the ideal generated mainly by the defining equations for the edge subring of a bipartite graph.

- Natural toric parametrization $\rightarrow$ `4ti2` computed a minimal generating set for the 4-node graph: $80,610$ binomials.
- Our parametrization $\rightarrow$ 77 minimal generators.
- Theorem + multi-grading $\rightarrow$ 10 essential 'pieces' of generators.

The hypergraph construction $\rightarrow$ a dynamic algorithms for generating squarefree Graver elements (Gross-P.-Stasi $'12^+$).

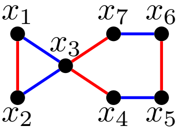




Degree bounds (Markov complexity)

Degree bounds (Markov complexity)

The graph case: a combinatorial criterion for I_G generated by quadrics –existence of special chords in G (Ohsugi-Hibi '00, Villarreal '01).

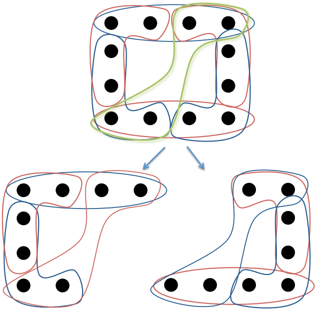


Degree bounds (Markov complexity)

The graph case: a combinatorial criterion for I_G generated by quadrics –existence of special chords in G (Ohsugi-Hibi '00, Villarreal '01).

Theorem (Gross-P. '12)

I_H is generated in degree at most d, for any $d \geq 2$, if and only if there exist appropriate splitting sets on balanced hypergraphs (binomial supports) on H.



Hypergraph with splitting set.

- The criterion is based on decomposable monomial walks, separators, and splitting sets.
- Non-existence of a splitting set $\Longrightarrow$ indispensable binomial in I_H.

A non-uniform hypergraph from statistics

Problem

Find the ideal of the first tangential variety $Tan((\mathbb{P}^1)^n)$.

- Sturmfels-Zwiernik '12? use cumulants to give a monomial parameterization of $\mathrm{Tan}((\mathbb{P}^1)^n)$.
- $\mathrm{Tan}((\mathbb{P}^1)^n)$ is associated to a class of hidden subset models
- ... and to context-specific independence models (Oeding '11).

Theorem (Gross-P. '12)

In cumulant coordinates, $Tan((\mathbb{P}^1)^n)$ is generated in degrees 2 and 3.

More precisely, for $H = (V, E)$ where $V = \{1, \dots, n\}$ and $E = \{e \subseteq V \ : 2 \leq |e| \leq 3\}$, I_H is generated by quadrics and cubics.

Problems

- The primitive walks in hypergraphs are clearly much more general than in the case of graphs.
- Nevertheless, we expect many of the other properties of the coordinate ring of I_H to have combinatorial interpretations.

1. Finding nice term orders and reduced Gröbner bases of I_H,
2. Characterize Cohen-Maculayness and normality of the coordinate ring and the corresponding polytope,
3. Relate known coloring-inspired properties of hypergraphs to various invariants of the toric ideal I_H.