Glimpses into recent developments in subword combinatorics

Markus A. Whiteland
Department of Computer Science
Loughborough University
m.a.whiteland@lboro.ac.uk

Based on joint works with

A. Renard, M. Rigo

30/06/2025
WORDS'25 @ Nancy

Outline

- Introduction/motivation
  - Subwords
  - Binomial coefficients of words
- q-deformations
  - Definitions
  - Properties
  - Deformations of known results
- Concluding remarks
  - Ongoing and future work

Introduction

Subwords

Subwords in the sciences

Binomial coefficient of words

Definition

v is a subword of u if v can be obtained by deleting letters from u.

- Example: a a b is a subword of a b b a b a
  - Note: sometimes called **scattered subwords** or **scattered factors.**
  - **Not the same as factors**; a factor is a subword but not always vice versa.

- In this talk: A^* is the set of finite words over alphabet A.
  - Always finite alphabets.
  - $a, b, c, \ldots$ indicate letters from A.
  - $u, v, w, \ldots$ indicate words from A^*.

- Practical problems call for reconstruction of data from subwords.
  - data transmission over lossy channels;
  - sequence kernels (in ML for text, biosequence, and time-series classification) as features capturing shared non-contiguous patterns;
  - protein sequence analysis in bioinformatics.

- Longest common subsequence, shortest common supersequence
  - with gap constraints (Adamson et al., WORDS'23)
- Language theory
  - Downward closures
- Logical theories for the subword relation
  - Halfon et al. (LICS'17)

- Simon's congurence
  - (near-)k-universality
  - Fleischmann et al.
- Palindromic subsequences
  - (Ago & Bašić WORDS'25)
- Jumbled subwords
  - (Fleischmann et al. DLT'25)
- binomial equivalence and complexity functions
  - (Rigo & co-authors since WORDS'13)
- multidimensional binomial coefficients
  - (Golafshan & Rigo WORDS'25)
- Binomial coefficients in generalised Pascal's triangle
  - Leroy, Rigo, & Stipulanti
- … (Apologies for missing many of the audience's works!)

Definition

For words $u, v \in A^*$, $\binom{u}{v}$ is the number of occurrences of v as a subword of u.

Example

$\binom{abbaba}{ab} = 4$

- Extends usual binomial coefficients over $\mathbb{N}$
  - unary free monoid → free monoids A^*.
  - $\binom{n}{m} = \binom{a^n}{a^m}$

- Kalashnik's reconstruction problem (1972):
  Determine the function f such that $f(n)$ is the least integer k such that any word u of length n has a unique vector $\left(\binom{u}{x}\right)_{x \in A^k}$.
  - Dudik & Schulman (2003)
  - Krasikov & Roditty (1997)
- Schützenberger's guessing game for reconstruction of an unknown word. Queries of the form: "What is $\binom{?}{x}$"
  - Fleischmann et al. (2021)
  - Richomme & Rosenfeld (STACS 2023)

Theorem (Eilenberg's p-group languages, 1976)

Let p be a prime. A language is a p-group language if and only if it is a Boolean combination of languages of the form $L_{v,r,p} := \left\{u \in A^* \mid \binom{u}{v} \equiv r \pmod{p}\right\}$.

For any word u of length n: $\sum_{x \in A^k} \binom{u}{x} = \binom{n}{k}$

Over $\mathbb{N}$

- Pascal's recurrence:

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$

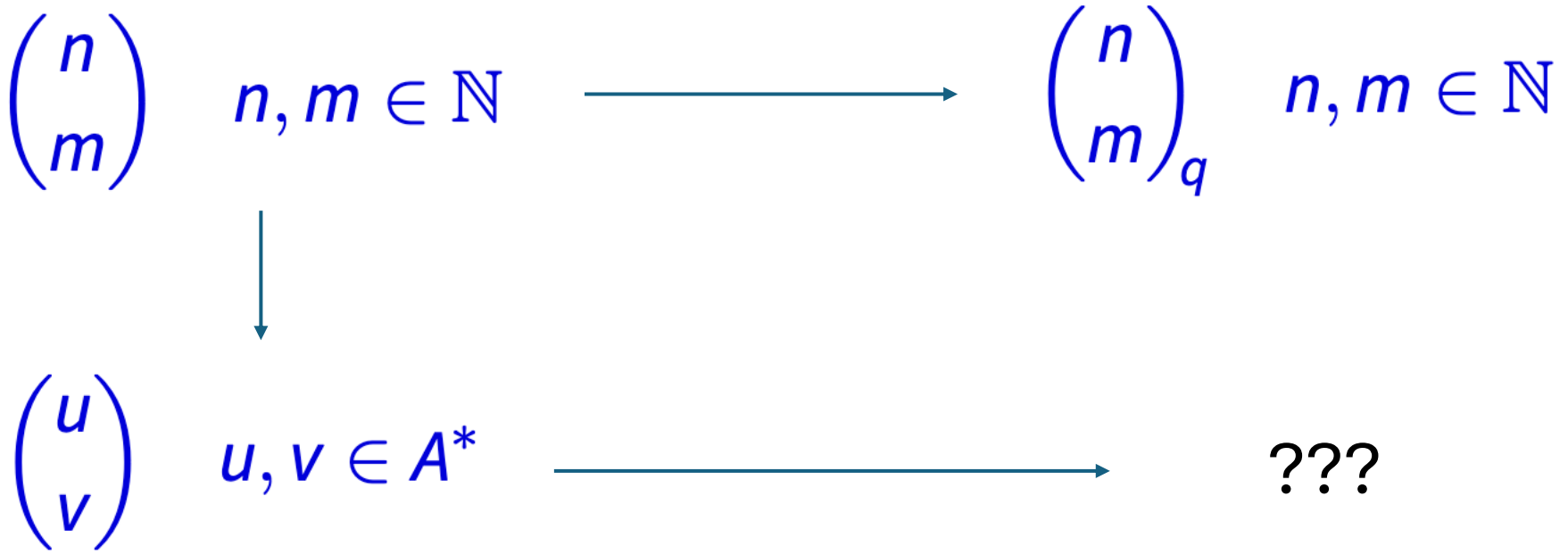
- Chu-Vandermonde identity:

$$\binom{n+m}{k} = \sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}$$

Over A^*

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

$$\binom{uw}{v} = \sum_{v_1 v_2 = v} \binom{u}{v_1}\binom{w}{v_2}$$



- Gaussian binomial coefficients

q-deformations

Gaussian binomial coefficients

q-binomial coefficients of words

Properties

q-analogs of theorems

- A q-deformation of an identity/expression is a generalisation involving a new parameter q that returns the original identity/expression in the limit $q \to 1$.

- E.g.,
  - q-natural numbers: $$[n]_q = \frac{1-q^n}{1-q} = 1 + q + \cdots + q^{n-1}$$
  - q-factorials: $$[n]_q! = [n]_q[n-1]_q \cdots [2]_q[1]_q$$
  - Gaussian binomial coefficients: $$\binom{n}{k}_q = \frac{[n]_q!}{[n-k]_q![k]_q!}$$
    - Always a polynomial.
    - $\binom{4}{2}_q = 1 + q + 2q^2 + q^3 + q^4$
    - Coefficient of q^d is the number of binary words with k many bs and d many inversions;
      - $abba$ and $baab$ have two inversion.

Binomial coefficients over $\mathbb{N}$

- Pascal's recursion
$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$
- Vandermonde identity
$$\binom{n+m}{k} = \sum_j \binom{n}{j}\binom{m}{k-j}$$
- …

Gaussian binomial coefficients

- q-Pascal recursion
$$\binom{n+1}{k+1}_q = q^{k+1}\binom{n}{k+1}_q + \binom{n}{k}_q$$
- q-Chu-Vandermonde identity
$$\binom{n+m}{k}_q = \sum_j q^{j(m-k+j)}\binom{n}{j}_q\binom{m}{k-j}_q$$
- …

Not only nice, but they have several interesting/useful interpretations

- $\binom{n}{k}_q$ is the number of dimension-k subspaces of $\mathbb{F}_q^n$ (q a prime power)
- $[q^r]\binom{n+k}{k}_q$ is the number of ways throwing r indistinguishable balls into k indistinguishable bins, each containing at most n balls.

q-deformations

Gaussian binomial coefficients

q-binomial coefficients of words

Properties

q-analogs of theorems

Based on: A. Renard, M. Rigo, W.: Introducing q-deformed binomial coefficients of words.
J Algebr Comb **61**:25 (2025). https://doi.org/10.1007/s10801-025-01384-9
A. Renard, M. Rigo, W.: q-Parikh matrices and q-deformed binomial coefficients of words,
Discrete Mathematics, **348**:5 (2025). https://doi.org/10.1016/j.disc.2024.114381

- Recall the q-Pascal recursion $$\binom{n+1}{k+1}_q = q^{k+1}\binom{n}{k+1}_q + \binom{n}{k}_q$$
- and the Pascal-like recursion $$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

Definition

Define the q-binomial coefficient $\binom{u}{v}_q$ recursively as

$$\binom{ua}{vb}_q = q^{|vb|}\binom{u}{vb}_q + \delta_{a,b}\binom{u}{v}_q; \quad \binom{u}{\varepsilon}_q = 1; \quad \binom{\varepsilon}{v}_q = 0 \quad (v \neq \varepsilon)$$

$$\binom{n+1}{k+1}_q = q^{k+1}\binom{n}{k+1}_q + \binom{n}{k}_q$$

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

$$\binom{ua}{vb}_q = q^{|vb|}\binom{u}{vb}_q + \delta_{a,b}\binom{u}{v}_q$$

- Examples:
  - $\binom{abbaba}{ab}_q = q^8 + q^7 + q^5 + q^2;$
  - $\binom{abbab}{ab}_q = q^6 + q^5 + q^3 + 1;$
  - $\binom{aabaa}{aa}_q = q^6 + q^4 + 2q^3 + q^2 + 1.$

- Clear from definition
  - $\lim_{q\to 1}\binom{u}{v}_q = \binom{u}{v};$
  - $\binom{a^n}{a^m}_q = \binom{n}{m}_q.$

Recall that $\binom{abbab}{ab} = 4$

- Is this the right definition?

$$\binom{ua}{vb}_q = q^{|vb|}\binom{u}{vb}_q + \delta_{a,b}\binom{u}{v}_q$$

  - Should be meaningful;
  - Should maintain “nice” properties/identities;
  - Have useful applications?

- $\binom{aabaa}{aa}_q = q^6 + q^4 + 2q^3 + q^2 + 1$

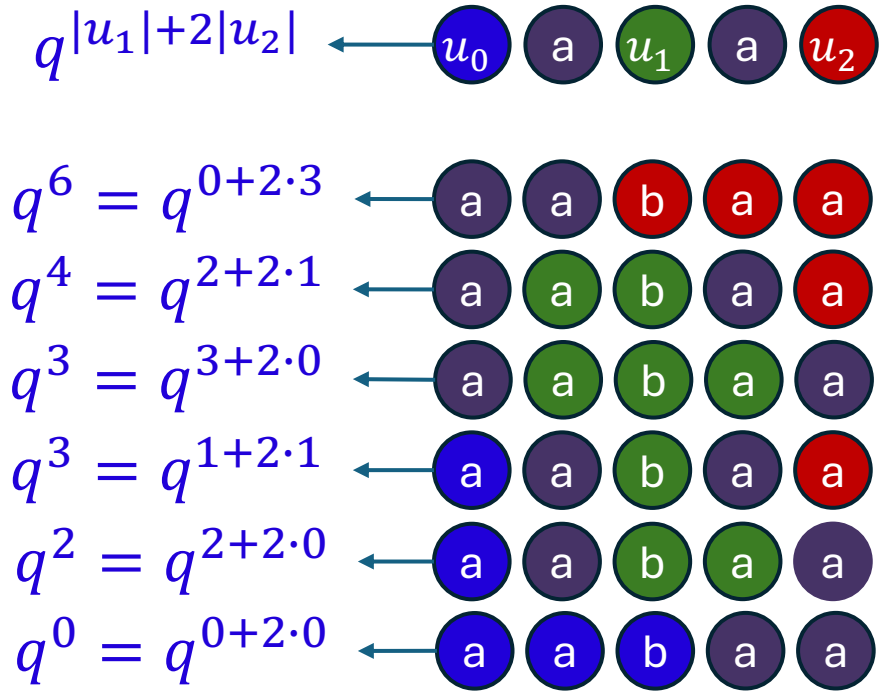
Where are these coming from?

Theorem (Renard, Rigo, W., 2025)

Let u be a word over A, $k \geq 0$, and $a_1, \ldots, a_k \in A$.

Then $$\binom{u}{a_1 \cdots a_k}_q = \sum_{\substack{u_0,u_1,\ldots,u_k \in A^* \\ u = u_0 a_1 \cdots u_{k-1} a_k u_k}} q^{\sum_{i=1}^{k} i|u_i|}.$$

- Each occurrence of the subword $a_1 \cdots a_k$ gives a monomial q^α



q-deformations

Gaussian binomial coefficients

q-binomial coefficients of words

Properties

q-analogs of theorems

Based on: A. Renard, M. Rigo, W.: Introducing q-deformed binomial coefficients of words.
J Algebr Comb **61**:25 (2025). https://doi.org/10.1007/s10801-025-01384-9
A. Renard, M. Rigo, W.: q-Parikh matrices and q-deformed binomial coefficients of words,
Discrete Mathematics, **348**:5 (2025). https://doi.org/10.1016/j.disc.2024.114381

Binomial coefficients over A^*

- Pascal-like recursion

$$\binom{ua}{vb} = \binom{u}{vb} + \delta_{a,b}\binom{u}{v}$$

- q-Vandermonde identity

$$\binom{n+m}{k}_q = \sum_j q^{j(m-k+j)} \binom{n}{j}_q \binom{m}{k-j}_q$$

- Vandermonde for words

$$\binom{uw}{v} = \sum_{v_1 v_2 = v} \binom{u}{v_1}\binom{w}{v_2}$$

- Other identities ...
- $\sum_{v \in A^k} \binom{u}{v} = \binom{|u|}{k}$

q-binomial coefficients over A^*

- Definition:

$$\binom{ua}{vb}_q = q^{|vb|}\binom{u}{vb}_q + \delta_{a,b}\binom{u}{v}_q$$

- q-Vandermonde for words

$$\binom{uw}{v}_q = \sum_{v=v_1v_2} q^{|v_1|(|w|-|v_2|)}\binom{u}{v_1}_q\binom{w}{v_2}_q$$

- Other q-identities ...
- $\sum_{v \in A^k} \binom{u}{v}_q = \binom{|u|}{k}_q$

Binomial coefficients over A^*

- (Generalised) Parikh matrices (à la Şerbănuţă 2004)
- Shuffle (formal series $\mathbb{N} \ll A^* \gg$) (à la Lothaire 1 (Sakarovitch & Simon))
- Infiltration (formal series $\mathbb{N} \ll A^* \gg$)

q-binomial coefficients over A^*

- q-Parikh matrices? ✓ (Initially studied by Eğecioğlu (2004))
- q-Shuffle? ✓
- q-infiltration? ⊖

q-deformations

Gaussian binomial coefficients

q-binomial coefficients of words

Properties

q-analogs of theorems

Based on: A. Renard, M. Rigo, W.: Introducing q-deformed binomial coefficients of words.
J Algebr Comb **61**:25 (2025). https://doi.org/10.1007/s10801-025-01384-9
A. Renard, M. Rigo, W.: q-Parikh matrices and q-deformed binomial coefficients of words,
Discrete Mathematics, **348**:5 (2025). https://doi.org/10.1016/j.disc.2024.114381

Definition

Recall that a language $L \subseteq A^*$ is ***recognised*** by a monoid M if there exist a subset $S \subset M$ and a monoid morphism $\varphi : A^* \to M$, such that $L = \varphi^{-1}(S)$.

A language is ***recognisable*** if it is recognised by a **finite** monoid.

A language is a p-group language if it is recognised by a p-group.

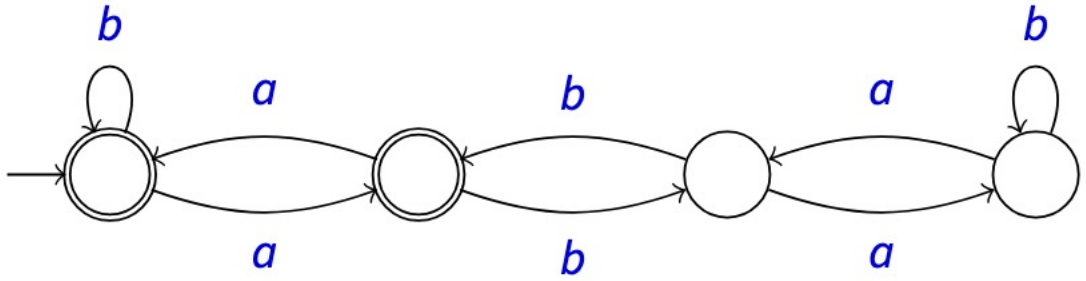
Recognisable languages = Regular languages

Theorem (Eilenberg's p-group languages, 1976)

Let p be a prime. A language is a p-group language if and only if it is a Boolean combination of languages of the form $L_{v,r,p} := \left\{ u \in A^* \mid \binom{u}{v} \equiv r \pmod{p} \right\}$.

r an element of $\mathbb{F}_p$

Example: $L_{ab,0,2}$



We consider languages of the form $L_{v,R,M} := \left\{ u \in A^* \mid \binom{u}{v}_q \equiv R \pmod{M} \right\}$, where $M \in \mathbb{F}_p[q]$ with $\deg(M) \geq 1$ and $R \in \mathbb{F}_p[q]$ with $\deg(R) < \deg(M)$.

- We follow Eilenberg's approach to define a congruence on A^* using q-binomial coefficients.

Definition ((u, M)-binomial equivalence relation)

Let $u \in A^+$ and $M \in \mathbb{F}_p[q]$ with $\deg(M) \geq 1$. Define

$$w_1 \sim_{u,M} w_2 \Leftrightarrow \forall v \in \mathrm{Fac}(u) : \binom{w_1}{v}_q \equiv \binom{w_2}{v}_q \pmod{M}.$$

- Problem: not always a congruence
- Solution: Consider congruences that refining $\sim_{u,M}$

Definition ((u, M)-binomial equivalence relation)

Let $u \in A^+$ and $M \in \mathbb{F}_p[q]$ with $\deg(M) \geq 1$. Define

$$w_1 \sim_{u,M} w_2 \Leftrightarrow \forall v \in \mathrm{Fac}(u) : \binom{w_1}{v}_q \equiv \binom{w_2}{v}_q \pmod{M}.$$

- If q is not invertible in $\mathbb{F}_p[q]/\langle M \rangle$, then $A^*/\equiv$ is **not** a group for any congruence $\equiv$ that refines $\sim_{u,M}$.

- Set $\equiv_{u,M}$ to be the coarsest congruence that refines $\sim_{u,M}$.
  If q is invertible in $\mathbb{F}_p[q]/\langle M \rangle$, then $A^*/\equiv_{u,M}$ is a group.

- The group has order that divides $\mathrm{per}(q) \cdot p^{|u|}$.

- Simple polynomial algebra: q is invertible in $\mathbb{F}_p[q]/\langle M \rangle$ and has $per(q)$ a power of p, then $M = t(q-1)^d$ for some non-zero t.

Theorem (Renard, Rigo, W., 2025)

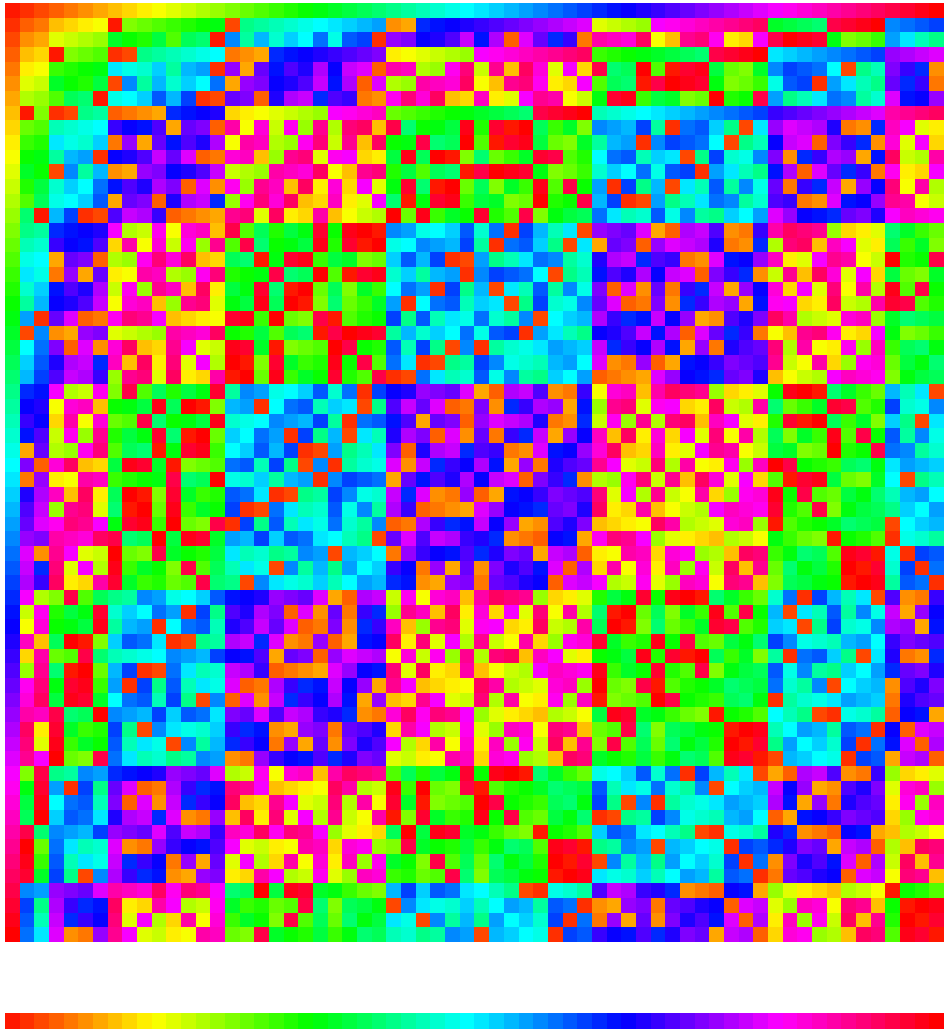
Let p be a prime, $d \geq 1$ an integer, and $t \in \mathbb{F}_p^*$ non-zero. A language is a p-group language if and only if it is a Boolean combination of languages of the form

$$L_{v,R} = \left\{ u \in A^* \mid \binom{u}{v}_q \equiv R \pmod{t(q-1)^d} \right\},$$

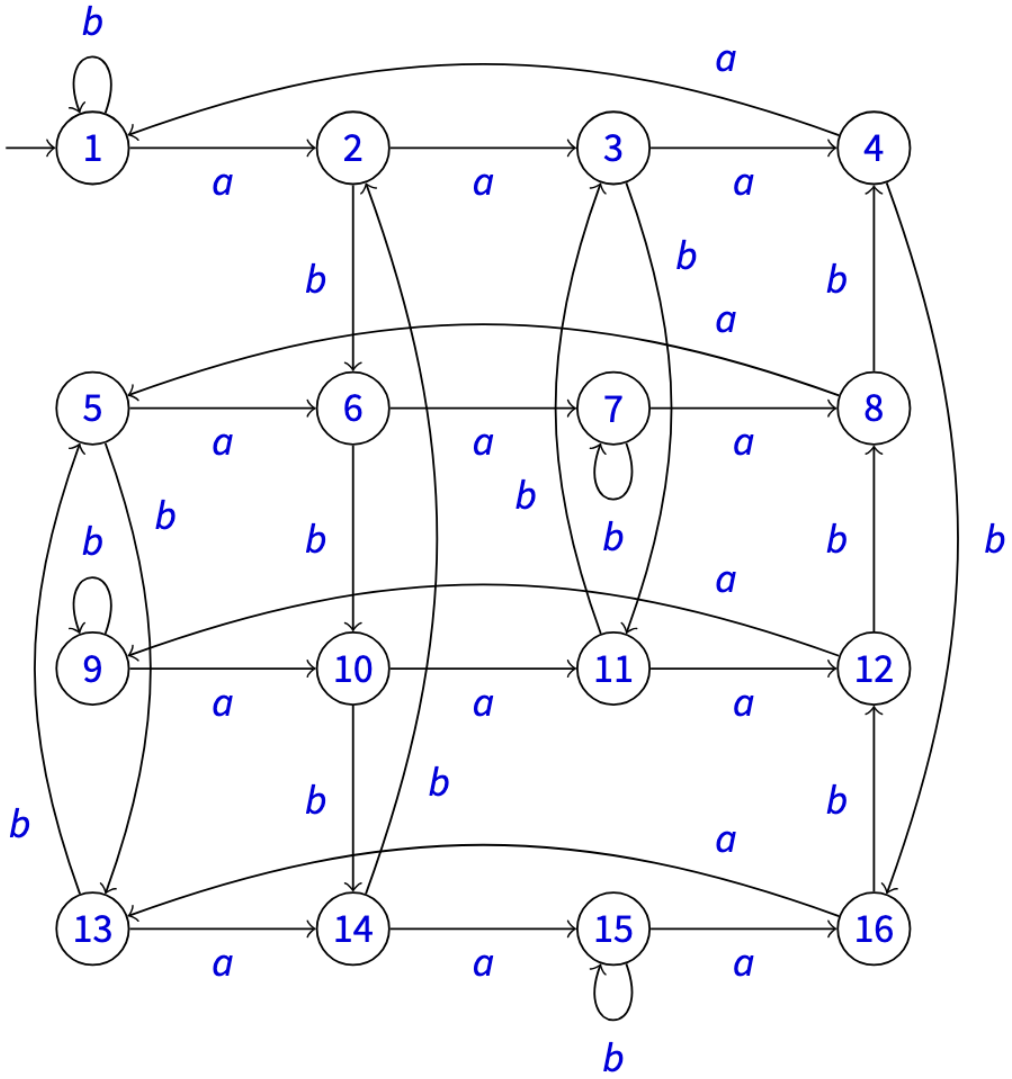
where $R \in \mathbb{F}_p[q]$ has $\deg(R) \leq d$.

Proof sketch: Let $\pi\colon A^* \to G := A^*/\equiv_{u,M}$, $x \mapsto [x]$. Take $S \subseteq G$ the classes $[y]$ for which $\binom{y}{v}_q \equiv R \ (mod\ M)$. Then $\pi^{-1}(S) = L_{v,R,t(q-1)^d}$.

Finally, Eilenberg's $L_{v,r,p}$ is a disjoint union of those $L_{v,R,t(q-1)^d}$, where $R(1) = r \ (mod\ p)$.



Group table of $\{a, b\}^*/\sim_{ab,q^2+1}$



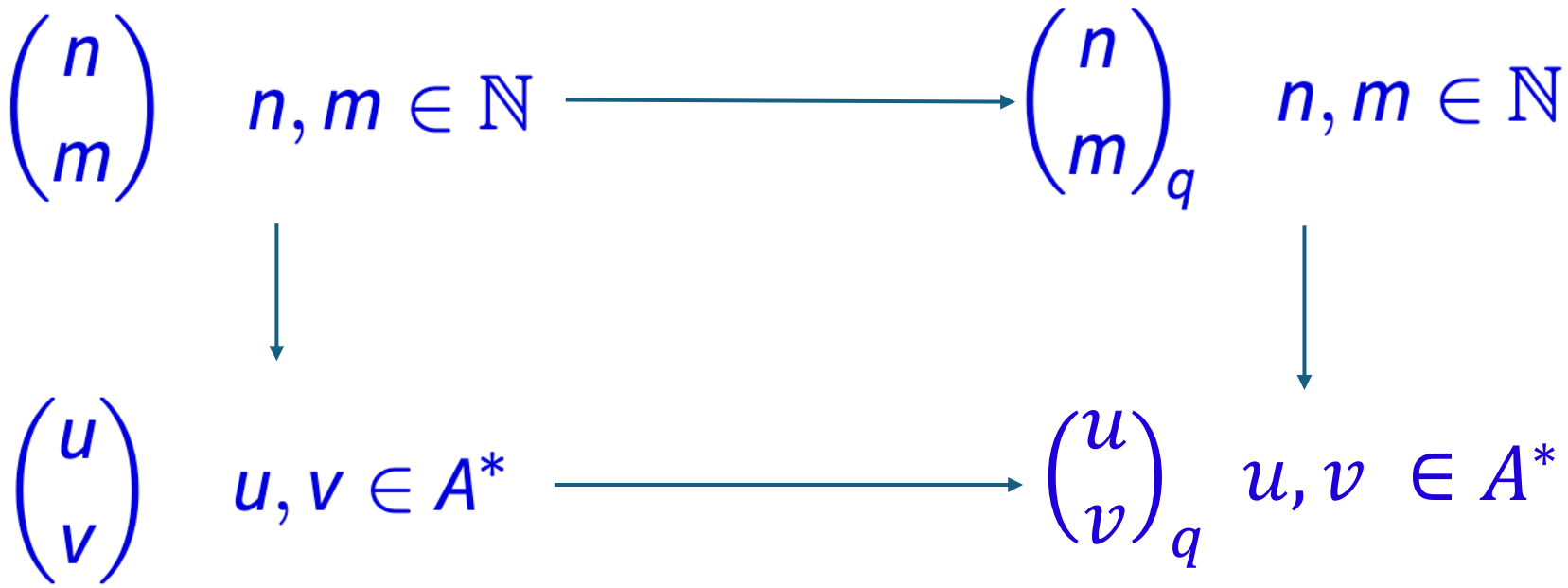
Automaton for L_{ab,R,q^2+1}

Conclusions

Recap: q-binomial coefficients of words

Ongoing work: q-binomial coefficients over free monoid?

Other directions: invitation



- Defined q-binomial coefficients for words
- Studied basic properties and found q-analog of Eilenberg's p-group languages.
- Conclusion: seems to be a "good" deformation
- A lot of open directions...

Definition ((u, M)-binomial equivalence relation)

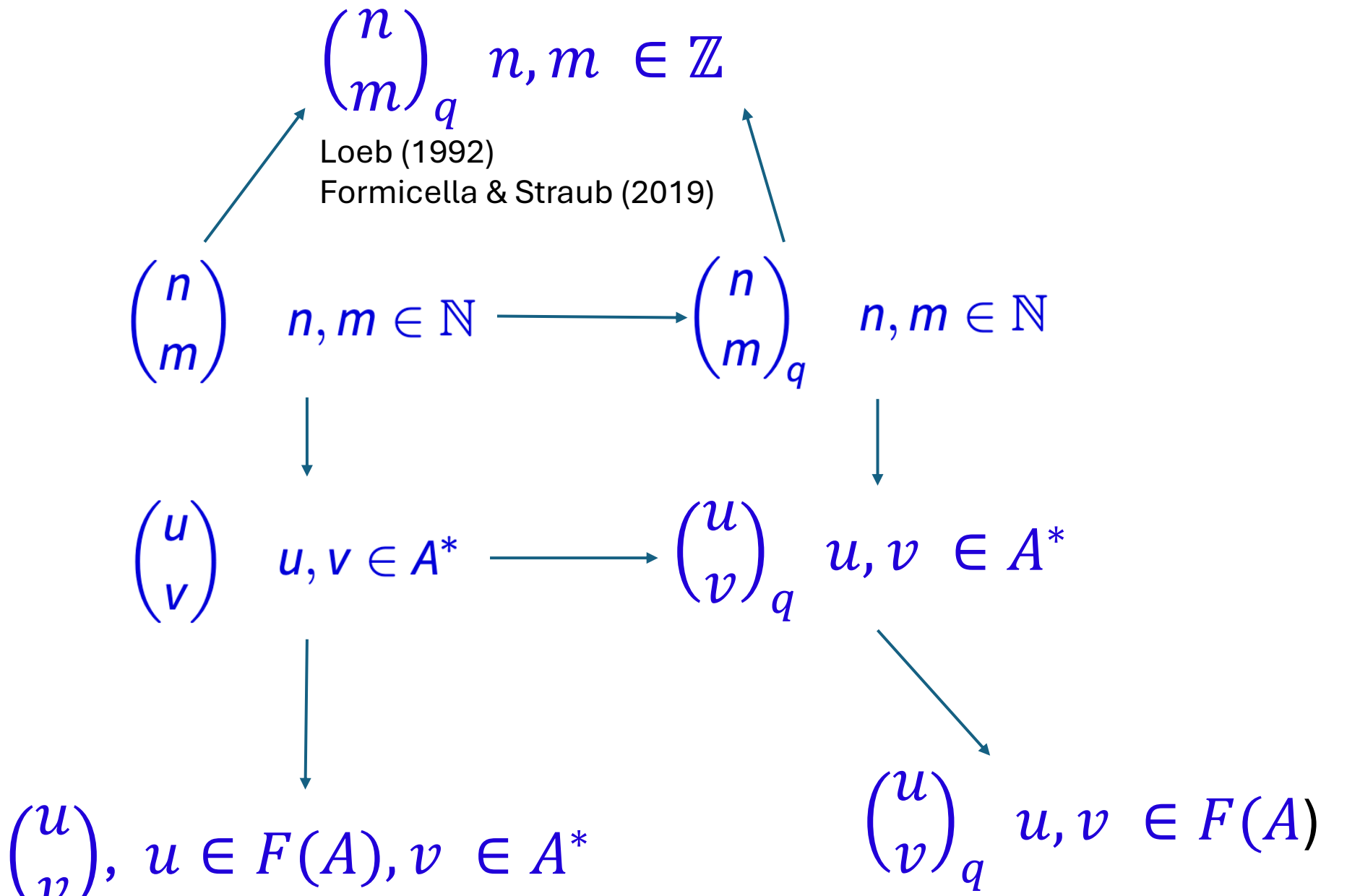
Let $u \in A^+$ and $M \in \mathbb{F}_p[q]$ with $\deg(M) \geq 1$. Define

$$w_1 \sim_{u,M} w_2 \Leftrightarrow \forall v \in \mathrm{Fac}(u) : \binom{w_1}{v}_q \equiv \binom{w_2}{v}_q \pmod{M}.$$

- If q is not invertible in $\mathbb{F}_p[q]/\langle M \rangle$, then $A^*/\equiv$ is **not** a group for any congruence $\equiv$ that refines $\sim_{u,M}$.

- Set $\equiv_{u,M}$ to be the coarsest congruence that refines $\sim_{u,M}$.
  If q is invertible in $\mathbb{F}_p[q]/\langle M \rangle$, then $A^*/\equiv_{u,M}$ is a group.
- The group has order that divides $\mathrm{per}(q) \cdot p^{|u|}$.

- What can be said about these languages?
  - Seem to be quite structured...


$\binom{n}{m}_q$ $n, m \in \mathbb{Z}$
Loeb (1992)
Formicella & Straub (2019)
$\binom{n}{m}$ $n, m \in \mathbb{N}$
$\binom{n}{m}_q$ $n, m \in \mathbb{N}$
$\binom{u}{v}$ $u, v \in A^*$
$\binom{u}{v}_q$ $u, v \in A^*$
$\binom{u}{v}$, $u \in F(A), v \in A^*$
Reutenauer (1993)
$\binom{u}{v}_q$ $u, v \in F(A)$
Together with J.-E. Pin (2025+)

- Gaussian binomial coefficients are products of cyclotomic polynomials.
  - Roots are roots of unity.
- This is not the case for q-binomial coefficients of words.
  - We have a conjecture on the language of (binary) words u such that for all x, $\binom{u}{x}_q$ is a product of cyclotomic polynomials and powers of q.
  - Any root is 0 or a root of unity.

- How large can the maximal-in-modulus root be?
  - Mild experiments suggest: < 2 in absolute value
- Combinatorial consequences from roots?

Thanks for your attention!

Happy to take any questions or comments!