Enumeration of Vertex Induced Subgraphs

Peter Tittmann (1)

Ilia Averbouch (2)

Johann A. Makowsky (2,*)

(1) Fachbereich Mathematik–Physik–Informatik
Hochschule Mittweida
Mittweida, Germany

(2) Faculty of Computer Science
Technion - Israel Institute of Technology
Haifa, Israel

Part of the Graph Polynomial Project

http://www.cs.technion.ac.il/~janos/RESEARCH/gp-homepage.html

(*) funded partially by the Israeli Science Foundation (2007-2010).

So many people today - and even professional scientists -
seem to me like someone has seen thousands of trees
but has never seen a forest.

A knowledge of the historical and philosophical background gives
that kind of independence from prejudices of his generation
from which most scientists are suffering.
This independence created by philosophical insight
is - in my opinion - the mark of distinction between
a mere artisan or specialist and a real seeker after truth.

A. Einstein

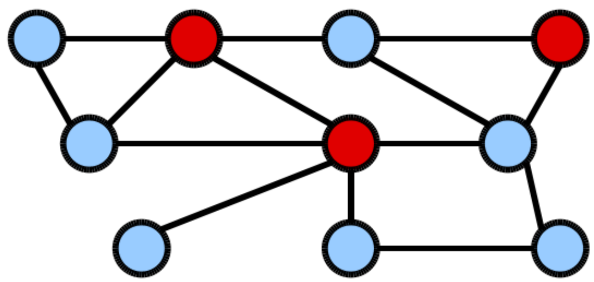
(from a letter to Robert Thornton, dated 7 December 1944, Einstein Archive 61-754),

Overview

- The subgraph component polynomial
- Recursive definition
- Universality property
- Connection to homomorphism functions
- Complexity issues
- Distinctive power and connection to other graph parameters

Background: social networks

Given a connected social network $G = (V, E)$



Question: How strong is the connection?

In other words: If some of the vertices randomly fail with probability p, how much connected components survive?

This leads to a new graph polynomial.

The Subgraph Component Polynomial

The Subgraph Component Polynomial

Definition 1

Let $G = (V, E)$ be a simple loop-free graph with $|V| = n$, and let $q_{ij}(G)$ denote the number of *induced subgraphs* of G with exactly i vertices and exactly j connected components:

$$q_{ij}(G) = |\{X \subseteq V : |X| = i \wedge k(G[X]) = j\}|,$$

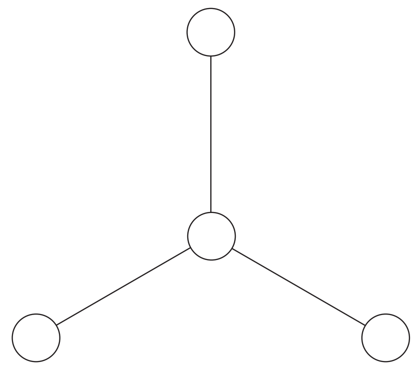
where $G[X]$ denotes the induced subgraph of G with vertex set X.

By convention, $q_{00} = 1$.

The *Subgraph Component Polynomial* is defined as the generating function

$$Q(G; x, y) = \sum_{i=0}^{n} \sum_{j=0}^{n} q_{ij}(G) x^i y^j \qquad (1)$$

The Subgraph Component Polynomial - Example



The star $K_{1,3}$ has the subgraph polynomial

$$Q\left(K_{1,3}; x, y\right) = 1 + 4xy + 3x^2y + 3x^3y + x^4y + 3x^2y^2 + x^3y^3.$$

The term $3x^2y^2$ tell us that there are 3 possibilities to select two vertices of G that are non-adjacent.

Substitution of 1 for y results in an univariate polynomial that is the ordinary generating function for all subsets of V, i.e. $Q(G; x, 1) = (1 + x)^n$.

The check list for new graph polynomials

Every time a new graph polynomial appears,

it is customary and natural to ask the following questions:

(i) Can it be presented as subset-expansion formula?
or better even: as an $\mathbf{MSOL}$-definable subset-expansion formula?

(ii) Does it satisfy some linear recurrence relation?

(iii) Is it definable as a partition function counting weighted homomorphisms?

(iv) How hard is it to compute?

(v) What is its connection to known graph polynomials?

(vi) and finally: **Is it really new?**

The check list, step by step

(i) It can be presented as an **MSOL**-definable subset-expansion formula, and is multiplicative!

(ii) Does it satisfy a linear recurrence relation?

(iii) Is it definable as a partition function?

(iv) How hard is it to compute?

(v) What is its connection to known graph polynomials?

(vi) and finally: **Is it really new?**

A subset expansion formula

Theorem 2 *The subgraph component polynomial can be presented as a vertex subset expansion* $\mathbf{MSOL}$*-formula.*

Instead of a summation over number of vertices i, let us summate over all the possible subsets of vertices $X \subseteq V$:

$$Q(G; x, y) = \sum_{X \subseteq V} x^{|X|} y^{k(G[X])}. \tag{2}$$

To express $k(G[X])$ we need an auxiliary order $\prec$ over the vertex set V. Then we count the "smallest" vertices in every connected component:

$$Conn(U) = (\forall W \subseteq U (\exists e = (u, v) \in E(u \in W \wedge v \in U \setminus W)))$$

$$First(U) = \{u : \forall W \subseteq U\, ((Conn(W) \wedge (u \in W)) \rightarrow (\forall v \in W (u \prec v)))\}$$

$$Q(G; x, y) = \sum_{X \subseteq V} \left(\prod_{v \in X} x \right) \left(\prod_{v \in First(X)} y \right) \tag{3}$$

Note that the result is independent on the order $\prec$.

Multiplicativity

Theorem 3 *Let* $G = G_1 \sqcup G_2$ *be disjoint union of the graphs* G_1 *and* G_2*. Then*

$$Q(G; x, y) = Q(G_1; x, y) \cdot Q(G_2; x, y) \tag{4}$$

Proof:
Every vertex subset $X = X_1 \sqcup X_2$ s.t. $X_1 = V(G_1) \cap X$ and $X_2 = V(G_2) \cap X$. Hense, $First(X) = First(X_1) \sqcup First(X_2)$. By (3) we have:

$$\begin{aligned} Q(G; x, y) &= \sum_{X \subseteq V} \left(\prod_{v \in X} x \right) \left(\prod_{v \in First(X)} y \right) = \\ &= \sum_{(X_1 \sqcup X_2) \subseteq V} \left(\prod_{v \in X_1} x \right) \left(\prod_{v \in X_2} x \right) \left(\prod_{v \in First(X_1)} y \right) \left(\prod_{v \in First(X_2)} y \right) = \\ &= Q(G_1; x, y) \cdot Q(G_2; x, y) \end{aligned}$$

Q.E.D.

The check list again

(i) It can be presented as $\mathbf{MSOL}$-definable subset-expansion formula!

(ii) It does satisfy a linear recurrence relation
and is universal for it!

(iii) Is it definable as a partition function?

(iv) How hard is it to compute?

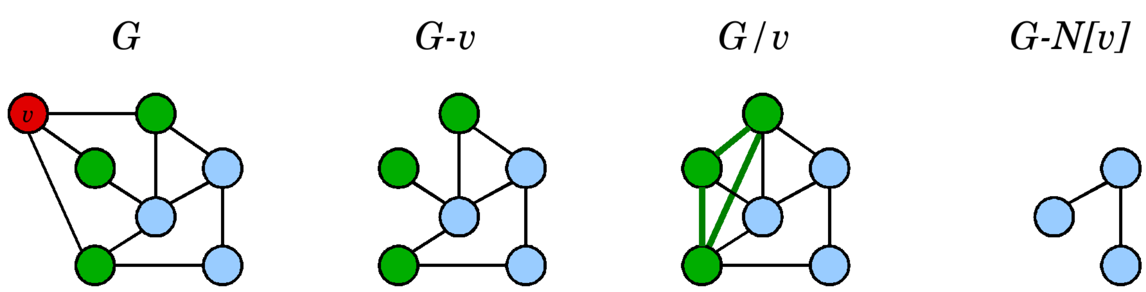
(v) What is its connection to known graph polynomials?

(vi) and finally: **Is it really new?**

Vertex elimination operations

Let $v \in V(G)$ be the vertex we want to remove:

- **Vertex deletion** $G - v$: induced subgraph of G with vertex set $V \setminus \{v\}$
- **Vertex contraction** G/v: the graph obtained from G by removing v and connecting all the vertices adjacent to v to clique.
- **Vertex extraction** $G - N[v]$: the graph obtained from G by removing v together with its neighborhood.



Decomposition formula, I

Theorem 4 *Let $G = (V, E)$ be a graph and $v \in V$. The subgraph component polynomial satisfies the decomposition formula*

$$Q(G; x, y) = Q(G - v; x, y) + xQ(G/v; x, y) + x(y - 1)Q(G - N[v]; x, y).$$

Proof:
We split all the subsets $X \subseteq V$ into three disjoint cases:

Case 1: $v \notin X$

Case 2: $v \in X$ but none of its neighbors in X

Case 3: v and at least one of its neighbors in X

Decomposition formula, II

Theorem 4 *Let $G = (V, E)$ be a graph and $v \in V$. The subgraph component polynomial satisfies the decomposition formula*

$$Q(G; x, y) = Q(G - v; x, y) + xQ(G/v; x, y) + x(y - 1)Q(G - N[v]; x, y).$$

Proof:
We split all the subsets $X \subseteq V$ into three disjoint cases:

Case 1:	$v \notin X$	$Q(G - v; x, y)$
Case 2:	$v \in X$ but none of its neighbors in X	
Case 3:	v and at least one of its neighbors in X	

Decomposition formula, III

Theorem 4 *Let $G = (V, E)$ be a graph and $v \in V$. The subgraph component polynomial satisfies the decomposition formula*

$$Q(G; x, y) = Q(G - v; x, y) + xQ(G/v; x, y) + x(y - 1)Q(G - N[v]; x, y).$$

Proof:
We split all the subsets $X \subseteq V$ into three disjoint cases:

Case 1:	$v \notin X$	$Q(G - v; x, y)$
Case 2:	$v \in X$ but none of its neighbors in X	$xy \cdot Q(G - N[v]; x, y)$
Case 3:	v and at least one of its neighbors in X	

Decomposition formula, IV

Theorem 4 *Let $G = (V, E)$ be a graph and $v \in V$. The subgraph component polynomial satisfies the decomposition formula*

$$Q(G; x, y) = Q(G - v; x, y) + xQ(G/v; x, y) + x(y - 1)Q(G - N[v]; x, y).$$

Proof:
We split all the subsets $X \subseteq V$ into three disjoint cases:

Case 1:	$v \notin X$	$Q(G - v; x, y)$
Case 2:	$v \in X$ but none of its neighbors in X	$xy \cdot Q(G - N[v]; x, y)$
Case 3:	v and at least one of its neighbors in X	$x \cdot (Q(G/v; x, y) - Q(G - N[v]; x, y))$

Decomposition formula, V

Theorem 4 *Let $G = (V, E)$ be a graph and $v \in V$. The subgraph component polynomial satisfies the decomposition formula*

$$Q(G; x, y) = Q(G - v; x, y) + xQ(G/v; x, y) + x(y - 1)Q(G - N[v]; x, y). \quad (5)$$

Proof:
We split all the subsets $X \subseteq V$ into three disjoint cases:

Case 1:	$v \notin X$	$Q(G - v; x, y)$
Case 2:	$v \in X$ but none of its neighbors in X	$xy \cdot Q(G - N[v]; x, y)$
Case 3:	v and at least one of its neighbors in X	$x \cdot (Q(G/v; x, y) - Q(G - N[v]; x, y))$

By (3) the theorem follows. Q.E.D.

Universality property

We call a graph polynomial $p(G)$ *universal* in class $\mathcal{P}$ if any other graph polynomial $r(G) \in \mathcal{P}$ can be reduced to $p(G)$.

Possible reductions are:

- A variable substitution: $r(G;\bar{x}) = p(G;\sigma(\bar{x}))$;
- Prefactoring: $r(G) = \tau(G)p(G)$;
- Graph transduction: $r(G) = p(T(G))$;
- Operation on coefficients: $[y_i]r(G;\bar{y}) = \rho_i([x_{i1}]p(G,\bar{x}),[x_{i2}]p(G,\bar{x}),\dots,[x_{ik}]p(G,\bar{x}))$;

Let $\mathcal{P}$ be a class of multiplicative graph polynomials that satisfy a linear recurrence relation with respect to vertex **deletion, contraction and extraction** operations above.

Question: is $Q(G;x,y)$ universal in $\mathcal{P}$?

$Q(G;x,y)$ is a universal vertex elimination polynomial

Theorem 5 *Every graph polynomial $p \in \mathcal{P}$,*
except for the Independent Set polynomial $I(G;x)$,
can be obtained from $Q(G;x,y)$ by simple variable substitution.

Proof:
Let us define the most general multiplicative generating function f satisfying a linear requrrence relation with respect to the three vertex elimination operations:

(a) (Multiplicativity) $f(G_1 \sqcup G_2) = f(G_1)\, f(G_2)$.

(b) (Recurrence relation) Let $\alpha, \beta, \gamma \in \mathbb{R}$ and let v be a vertex of G, then

$$f(G) = \alpha f(G - v) + \beta f(G - N[v]) + \gamma f(G/v)\,. \qquad (6)$$

(c) (Initial condition) There exists $\delta \in \mathbb{R}$ such that $f(\emptyset) = \delta$ for the null graph $\emptyset = (\emptyset, \emptyset)$.

(d) (Initial condition) There exists $\varepsilon \in \mathbb{R}$ such that $f(E_1) = \varepsilon$ for a graph $E_1 = (\{v\}, \emptyset)$ consisting of one vertex.

$Q(G;x,y)$ is a universal vertex elimination polynomial

We exploit the fact that $f(G)$ is a *graph invariant*:

- Disjoint union with $\emptyset$: $f(G \sqcup \emptyset) = f(G) \Rightarrow \delta = 1$
- $f(\emptyset) = f(E_1 - v) = f(E_1/v) = f(E_1 - N[v]) \Rightarrow \varepsilon = (\alpha+\beta+\gamma)$
- $f(P_3)$ does not depend on the order of decomposition.

$$\left.\begin{array}{l} f(P_3) = (\alpha+\gamma)^2(\alpha+\beta+\gamma) + \beta(\alpha+\gamma) + \beta(\alpha+\beta+\gamma) \\ f(P_3) = \alpha(\alpha+\beta+\gamma)^2 + \beta + \gamma(\alpha+\gamma)(\alpha+\beta+\gamma) + \beta\gamma \end{array}\right\} \quad \begin{array}{l} \alpha = 1 \quad or \\ \beta = 0 \quad or \\ (\alpha+\beta+\gamma) = 1 \end{array}$$

- $(\alpha+\beta+\gamma) = 1$ leads to a trivial $f(G) = 1$ for any G;
- $(\beta = 0)$ leads to a trivial $f(G) = (\alpha+\gamma)^{|V|}$ for any G;
- $\alpha = 1, \gamma = 0$ leads to $f(G) = I(G;\beta)$, the Independent Set polynomial;
- $\alpha = 1, \gamma \neq 0$ leads to $f(G) = Q(G;\gamma,\frac{\beta}{\gamma}+1)$

The check list again

(i) It can be presented as $\mathbf{MSOL}$-definable subset-expansion formula!

(ii) It does satisfy a linear recurrence relation and is universal for it!

(iii) It is definable as a partition function!

(iv) How hard is it to compute?

(v) What is its connection to known graph polynomials?

(vi) and finally: **Is it really new?**

Partition functions

Definition 6 *Counting weighted graph homomorphisms.*

Let $H = (V_H, E_H)$ be a labeled graph, let α to assign weights to its vertices, and let β to assigh weights to its edges. The *partition function* $Z_H(G)$ counts weighted graph homomorphisms from G to H:

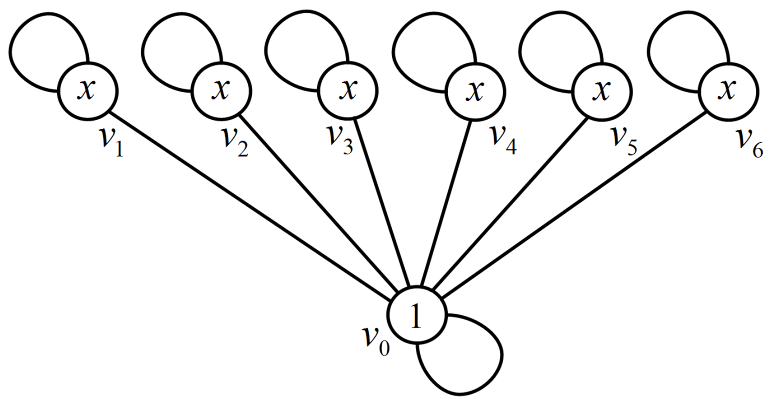
$$Z_H(G) = \sum_{\substack{h\,:\,V \mapsto V_H \\ homomorphism}} \prod_{v \in V} \alpha(h(v)) \prod_{(u,v) \in E} \beta(h(u), h(v))$$

Question: can $Q(G; x, y)$ be presented as a partition function?

Partition functions

Let H be a star $Star_n$ with all loops and central vertex v_0, let the weight of all the edges be $\beta = 1$, and weight of the vertices as follows:

$$\alpha(v) = \begin{cases} 1 & if\ v = v_0 \\ x & otherwise \end{cases}$$



Theorem 7 *For all nonnegative integers $n \in \mathbb{Z}_+$ and all real $x \in \mathbb{R}$,*

$$Q(G; x, n) = Z_{Star_n}(G),$$

where $Z_{Star_n}(G)$ is a partition function asociated with $Star_n$, α and β above.

The check list again

(i) It can be presented as **MSOL**-definable subset-expansion formula!

(ii) It does satisfy a linear recurrence relation and is universal for it!

(iii) It is definable as a partition function!

(iv) It is mostly hard to compute.

It is easy on graphs of fixed tree width and clique width.

(v) What is its connection to known graph polynomials?

(vi) and finally: **Is it really new?**

Complexity: Hardness

Theorem 8 Complexity of evaluation*:*
For every point $(x, y) \in \mathbb{Q}^2$*,*
except for the lines $xy = 0$*,* $y = 1$*,* $x = -1$ *and* $x = -2$*,*
the evaluation of $Q(G; x, y)$ *for an input graph* G *is* $\sharp\mathbf{P}$*-hard.*

Proof:

Follows from Theorem 10 and Theorem (Hoffmann 2008):
"For every point $(x, y, z) \in \mathbb{Q}^3$*, except possibly for the subsets* $x = 0$*,* $z = -xy$*,* $(x, z) \in \{(1, 0), (2, 0)\}$ *and* $y \in \{-2, -1, 0\}$*, the evaluation of* $\xi(G; x, y, z)$ *for an input graph* G *is* $\sharp\mathbf{P}$*-hard"* Q.E.D.

Complexity: Fixed Parameter Tractable FPT

However,if the input graph is of bounded tree-width or of bounded clique-width, we can use Theorem 2, and apply general complexity meta-theorems (Courcelle, Makowsky, Rotics):

Proposition 9

$Q(G; x, y)$ is polynomial time computable on graph classes $\mathcal{K}$ for

(i) $\mathcal{K}$ the class of graphs of tree-width at most k, and

(ii) $\mathcal{K}$ the class of graphs of clique-width at most k,

where the exponent of the run time is independent of k. In other words it is in FPT

The check list again

(i) It can be presented as **MSOL**-definable subset-expansion formula!

(ii) It does satisfy a linear recurrence relation and is universal for it!

(iii) It is definable as a partition function!

(iv) It is mostly hard to compute.

It is easy on graphs of fixed tree width and clique width.

(v) There are connections to known graph polynomials.

But in many ways it behaves differently.

(vi) and finally: **Is it really new?**

Universal Edge Elimination Polynomial

The universal edge elimination polynomial $\xi(G; x, y, z)$ is defined as follows:

$$\begin{aligned} \xi(G; x, y, z) &= \xi(G - e; x, y, z) + y\xi(G/e; x, y, z) + z\xi(G \dagger e; x, y, z) \\ \xi(G_1 \sqcup G_2; x, y, z) &= \xi(G_1; x, y, z)\xi(G_2; x, y, z) \\ \xi(E_1; x, y, z) &= x \\ \xi(\emptyset) &= 1 \end{aligned} \tag{7}$$

where

Edge deletion: We denote by $G - e$ the graph obtained from G by simply removing the edge e.

Edge extraction: We denote by $G \dagger e$ the graph induced by $V \setminus \{u, v\}$ provided $e = \{u, v\}$. Note that this operation removes also all the edges adjacent to e.

Edge contraction: We denote by G/e the graph obtained from G by unifying the endpoints of e.

Question: how $Q(G; x, y)$ is related to $\xi(G; x, y, z)$?

Connection to $\xi(G; x, y, z)$

Theorem 10 *Let $G = (V, E)$ be a graph. Let $L(G) = (V_e, E_e)$ denote the line graph of G. Then the following equation holds:*

$$\xi(G; 1, x, x(y-1)) = Q(L(G); x, y)$$

Proof:
Connection between edge elimination and vertex elimination operations:

$$\begin{aligned} L(G-e) &= L(G) - v_e \\ L(G/e) &= L(G)/v_e \\ L(G \dagger e) &= L(G) - N[v_e] \end{aligned}$$

Initial conditions

$$\begin{aligned} G \in \{\emptyset, E_1\} \Rightarrow &\quad L(G) = \emptyset, \quad \xi(G; 1, x, x(y-1)) = 1 = Q(\emptyset) \\ G \in \{P_2, Loop_1\} \Rightarrow &\quad L(G) = E_1, \quad \xi(G; 1, x, x(y-1)) = 1 + xy = Q(E_1) \end{aligned}$$

Multiplicativity: $L(G_1 \sqcup G_2) = L(G_1) \sqcup L(G_2)$

The theorem follows by induction by the number of edges. Q.E.D.

Distinctive power, I

We say that a graph polynomial $p(G)$ *determines* graph invariant $r(G)$ if for every pair of graphs G_1 and G_2

$$p(G_1) = p(G_2) \Rightarrow r(G_1) = r(G_2)$$

Question:

Is $Q(G; x, y)$ determined by some known graph polynomial?

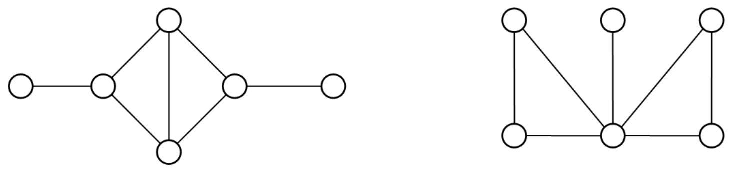
Distinctive power, II

We compare the distinctive power of $Q(G; x, y)$
to the known graph polynomials:

- The characteristic polynomial $p(G; x)$;
- The matching polynomial $m(G; x)$;
- The Tutte polynomial $T(G; x, y)$;
- The bivariate chromatic polynomial $P(G; x, y)$;

Distinctive power, III

Proposition 11 $Q(G;x,y)$ *is not determined by the characteristic polynomial* $p(G;x)$*.*



Graphs having the same $p(G;x)$ but different $Q(G;x,y)$

Proposition 12 $Q(G;x,y)$ *is not determined by the matching polynomial* $m(G;x)$*.*

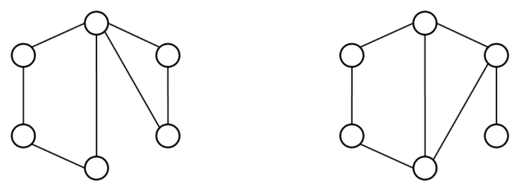
Graphs having the same $m(G;x)$ but different $Q(G;x,y)$

Distinctive power, IV

The bivariate chromatic polynomial $P(G, x, y)$
(K.Dohmen, A.Pönitz and the first author)
is defined as follows:

for two integers $x \geq y$, $P(G, x, y)$ is the number of colorings of G by x colors, y of them are proper.

Proposition 13 *$Q(G; x, y)$ is not determined by the bivariate chromatic polynomial $P(G, x, y)$.*



Graphs having the same $P(G; x, y)$ but different $Q(G; x, y)$

Proposition 14 *$Q(G; x, y)$ is not determined by the Tutte polynomial $T(G; x, y)$.*

Indeed, the Tutte polynomial does not distinguish between trees of the same size, whereas $Q(G; x, y)$ does for trees up to 9 vertices.

The check list again

(i) It can be presented as **MSOL**-definable subset-expansion formula!

(ii) It does satisfy a linear recurrence relation and is universal for it!

(iii) It is definable as a partition function!

(iv) It is mostly hard to compute.
It is easy on graphs of fixed tree width and clique width.

(v) There are connections to known graph polynomials.

But in many ways it behaves differently.

(vi) **It seems to be really new!**

Conclusions

- We have defined a new graph polynomial $Q(G;x,y)$ that is motivated by the connectivity features of social networks
- We have shown that this polynomial can be defined statically as a subset-expansion formula, recursively as a linear recurrence relation, or as a partition function
- We have shown its universality property with respect to the class of graph polynomials that satisfy a linear recurrence relation based on three vertex elimination operations
- We have analyzed the computational complexity of $Q(G;x,y)$.
- We have found the connection between $Q(G;x,y)$ and the universal edge elimination polynomial
- We have shown that $Q(G;x,y)$ is NOT determined by various known graph polynomials

Open questions

- We know the connection between $\xi(G)$ and $Q(L(G))$.
  However, we do not know whether $\xi(G)$ determines $Q(G)$.

  Question: Does $\xi(G)$ determine $Q(G)$?

- We know that $Q(G; x, y)$ is not determined
  by various known graph polynomials.

  Question: does $Q(G; x, y)$ determine some known graph polynomial?

- We know that $Q(G; x, y)$ is $\sharp\mathbf{P}$-hard to evaluate at $x, y \in (Q)$, except for the lines $xy = 0$, $y = 1$, $x \in \{-1, -2\}$.
  We know that it is easy to evaluate at the first two lines.

  Question: is $Q(G; x, y)$ hard to evaluate at $x \in -1, -2$?

Thank you for your attention!