79/24

Strong Independence of Graphcopy Functions

PAUL ERDÖS, LÁSZLÓ LOVÁSZ,

and

JOEL SPENCER

Let H be a finite graph on v vertices. We define a function c_H, with domain the set of all finite graphs, by letting $c_H(G)$ denote the fraction of subgraphs of G on v vertices isomorphic to H. Our primary aim is to investigate the behavior of the functions c_H with respect to each other. We show that the c_H, where H is restricted to be connected, are independent in a strong sense. We also show that, in an asymptotic sense, the c_H, H disconnected, may be expressed in terms of the c_H, H connected.

In 1932, Whitney [1] proved that the functions c_H, H connected, were algebraically independent. Our results may be considered an extension of this work.

Notations and Conventions

All graphs G shall be finite, without loops of multiple edges. $\mathscr{G}$ denotes the family of all finite graphs. $V(G)$, $E(G)$ denote the vertex and edge sets of G.

ISBN 0-12-114350-3

For purposes of counting all graphs may be considered labelled. A map

$$\psi: V(H) \to V(G)$$

is called a *homomorphism* if $\{x, y\} \in E(H)$ implies $\{\psi_x, \psi_y\} \in E(G)$. If furthermore $x \neq y$ implies $\psi_x \neq \psi_y$ then ψ is a *monomorphism*. For $W \subseteq V(G)$, the restriction of G to W, denoted by $G|_W$, is that graph with vertex set W and $\{x, y\} \in E(G|_W)$ iff $\{x, y\} \in E(G)$, $x, y \in W$. Let I_s, K_s denote the empty and complete graphs respectively on s elements.

Let $H, G \in \mathscr{G}$, $|V(H)| = t$, $|V(G)| = n$. Define

$$\begin{aligned}
A_H(G) &= |\{\psi: V(H) \to V(G), \text{homomorphism}\}|,\\
B_H(G) &= |\{\psi: V(H) \to V(G), \text{monomorphism}\}|,\\
C_H(G) &= |\{W \subseteq V(G): |W| = t, G|_W \cong H\}|,\\
a_H(G) &= A_H(G)/n^t,\\
b_H(G) &= B_H(G)/(n)_t \qquad [(n)_t = n(n-1)\cdots(n-t+1)],\\
c_H(G) &= C_H(G)/\tbinom{n}{t}.
\end{aligned}$$

The lower case functions give the fraction of homomorphism, monomorphism, and copies, respectively.

Throughout this paper let k denote a fixed integer $k \geq 3$. Let $H_1, \ldots, H_m$ be all connected graphs (up to isomorphism) with $|V(H_i)| \leq k$. Let A_i, B_i, C_i, a_i, b_i, c_i denote the functions $A_{H_i}, \ldots, c_{H_i}$ for convenience. Define vector valued functions $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ by

$$\begin{aligned}
\mathbf{a}(G) &= (a_1(G), \ldots, a_m(G)),\\
\mathbf{b}(G) &= (b_1(G), \ldots, b_m(G)),\\
\mathbf{c}(G) &= (c_1(G), \ldots, c_m(G)).
\end{aligned}$$

Our object is to study the possible values for $\mathbf{a}(G)$, $\mathbf{b}(G)$, $\mathbf{c}(G)$. We wish, however, to avoid exceptional values taken by small G.

Definition S_a, S_b, S_c are defined as the sets of limit points of $\{\mathbf{a}(G)\}$, $\{\mathbf{b}(G)\}$, $\{\mathbf{c}(G)\}$, respectively. More precisely, for $x = a, b, c$

$$S_x = \{\mathbf{v} \in R^m; \exists \text{ sequence } G_n, |V(G_n)| \to \infty, \mathbf{x}(G_n) \to \mathbf{v}\}.$$

Example 1 Let $k = 3$. Let H_1 be the 2-point, 1-edge graph; H_2 the 3-point 2-edge (vee) graph, H_3 the 3-point 3-edge (triangle) graph. Then

$$\mathbf{c}(H_2) = (\tfrac{2}{3}, 1, 0) \notin S_c.$$

We show later that $S_a = S_b$ and that S_c is a nonsingular linear transformation of S_a. Though the set S_c has perhaps the greatest natural interest, we shall prove theorems for S_a, where the technical problems are minimal.

Three Constructions

Let H be a graph on vertex set $V(H) = \{1, \ldots, t\}$. Let $x_1, \ldots, x_{t+1} \geq 0$, integral. We define a new graph $H^* = H(x_1, \ldots, x_t : x_{t+1})$ as follows:

$$V(H^*) = S_1 \cup \cdots \cup S_t \cup S_{t+1},$$

where S_i are disjoint sets $|S_i| = x_i$

$$E(H^*) = \{\{\alpha, \beta\} : \alpha \in S_i, \beta \in S_j, \{i, j\} \in E(H)\}.$$

Intuitively, we have blown up the ith vertex into an independent set of size x_i and added x_{t+1} isolated points.

Notation $H(x_1, \ldots, x_t) = H(x_1, \ldots, x_t : 0)$.

$\phi: H(x_1, \ldots, x_i) \to H$ is the canonical homomorphism defined by $\phi(\alpha) = i$ iff $\alpha \in S_i$,

$$sH = H(x_1, \ldots, x_t) \quad \text{where} \quad x_1 = \cdots = x_t = s,$$

$\prod_{i=1}^{t} G_i$ is the graph consisting of vertex disjoint copies of G_i,

$\prod_{i=1}^{t} a_i G_i$ is the graph consisting of vertex disjoint copies of $a_i G_i$.

Lemma 1 *If G has no isolated points*

$$A_G(H(x_1, \ldots, x_t : x_{t+1})) = \sum_{\psi} \prod_{i \in V(G)} x_{\psi(i)},$$

where ψ ranges over all homomorphisms $\psi: G \to H$. Also

$$a_G(H(x_1, \ldots, x_t : x_{t+1})) = \sum_{\psi} \prod_{i \in V(G)} P_{\psi(i)},$$

where we define $P_j = x_j / \sum_{i=1}^{t+1} x_i$.

Proof As G has no isolated points any homomorphism is into $H(x_1, \ldots, x_t)$. For each homomorphism $\lambda: G \to H(x_1, \ldots, x_t)$, $\psi = \phi\lambda$ is a homomorphism and for each ψ there are precisely $\prod_{i \in V(G)} x_{\psi_i}$ homomorphisms λ with $\psi = \phi\lambda$. The second equality follows from division. ■

Lemma 2

$$A_H(sG) = A_H(G) s^{|V(H)|},$$

$$a_H(sG) = a_H(G).$$

Lemma 2 is only a special case of Lemma 1.

Corollary 1 $\mathbf{a}(G) \in S_a$ *for all G.*

Proof Let $G_n = nG$ in the definition of S_a. ■

Example 1 shows the Corollary 1 does not hold for S_c; nor does it hold for S_b.

Lemma 3 *If H is connected*

$$A_H\left(\sum_{i=1}^{t} a_i G_i\right) = \sum_{i=1}^{t} A_H(G_i)a_i^{|H|},$$

$$a_H\left(\sum_{i=1}^{t} a_i G_i\right) = \sum_{i=1}^{t} a_H(G_i)p_i^{|H|},$$

where

$$p_i = a_i|V(G_i)| \Big/ \sum_{j=1}^{t} a_j|V(G_i)|.$$

Proof The first formula follows from Lemma 2 and the observation that the range of any homomorphism ψ shall be connected and hence lie in some $a_i G_i$. The second formula follows from division. ■

Let

$$G = \sum_{i=1}^{t} a_i G_i + I_{a_{t+1}}$$

and set $p_i = a_i|V(G_i)|/|V(G)|$ for $1 \leq i \leq t$. A simple calculation gives

Lemma 4 *For H connected and G given as above*

$$a_H(G) = \sum_{i=1}^{t} a_H(G_i)p_i^{|H_i|}.$$

Dimensions

Now we are able to state our main result.

Theorem 1 *There exist* $\mathbf{z} \in R^m$, $\varepsilon > 0$, *so that* $B(\mathbf{z}, \varepsilon) \subseteq S_a$. *Here* $B(\mathbf{z}, \varepsilon)$ *is the ball of radius* ε *about* $\mathbf{z}$.

We require a preliminary lemma.

Lemma 5

$$\{\mathbf{a}(G)\}_{G \in \mathscr{G}} \text{ span } R^m \text{ (as a vector space)}.$$

Proof We use an indirect argument. If the lemma is false there exist $c_1, \ldots, c_m$ not all zero, so that

$$\sum_{i=1}^{m} c_i a_i(G) = 0 \qquad \text{for all} \quad G \in \mathscr{G}.$$

From $\{H_i: c_i \neq 0\}$ select from among the H with any particular number v of vertices an H with the minimal number of edges. Let $G = H(x_1, \ldots, x_v: x_{v+1})$. From Lemma 1, $a_i(G)$ is a polynomial in $p_1, \ldots, p_v$ where $p_i = x_i/|V(G)|$. The coefficient of $p_1 \cdots p_v$ in $a_i(G)$ is the number of *bijective* homomorphisms $\psi: H_i \to H$. By the minimality of H this coefficient is nonzero *iff* $H = H_i$. Thus $\sum c_i a_i(G)$ is not the zero polynomial—but then there exist rational values $p_1, \ldots, p_v \geq 0, \sum_{i=1}^{v} p_i < 1$ for which the polynomial is nonzero. We may find $x_1, \ldots, x_{v+1} \in Z$ yielding these p_i, contradicting our assumption. ■

Proof of Theorem 1 Let $G_1, \ldots, G_m$ be such that $\{\mathbf{a}(G_i)\}$ span R^m. Set

$$a_j(G_i) = a_{ij}, \qquad 1 \leq i, j \leq n$$

so that the matrix $[a_{ij}]$ is nonsingular. Let $p_1, \ldots, p_m, p_{m+1} \geq 0$ and rational, with

$$\sum_{i=1}^{m+1} p_i = 1.$$

Let $D \in Z, D > 0$ so that all $Dp_i/|V(G_i)| \in Z$. Set

$$G = \sum_{i=1}^{m} \left[\frac{(Dp_i)}{|V(G_i)|} \right] G_i + I_{Dp_{m+1}}.$$

Then, by Lemma 4,

$$a_j(G_i) = \sum_{i=1}^{m} a_{ij} p_i^{|H_j|}, \qquad 1 \leq j \leq m. \tag{$\bigstar$}$$

We consider ($\bigstar$) as a map $\Psi: R^m \to R^m$ transforming coordinates $p_1, \ldots, p_m$ to $\eta_1, \ldots, \eta_m$. G is defined for all $0 \leq p_1, \ldots, p_m \leq m^{-1}$. Then

$$S_a \supseteq \{\Psi(p_1, \ldots, p_m): 0 \leq p_1, \ldots, p_m \leq m^{-1}, p_i \in Q\}.$$

Since Ψ is continuous and S_a closed

$$S_a \supseteq \{\Psi(p_1, \ldots, p_m): 0 \leq p_1, \ldots, p_m \leq m^{-1}\}.$$

Now we calculate

$$\frac{\partial \eta_j}{\partial p_i} = a_{ij} |H_j| p_i^{|H_j|-1}$$

and $Jac(\Psi)$ is a polynomial in $p_1, \ldots, p_m$. At $p_1 = \cdots = p_m = 1$

$$Jac(\Psi) = \det[a_{ij}|H_j|] = \left[\prod_{j=1}^{m} |H_j|\right] \det[a_{ij}] \neq 0,$$

so $Jac(\Psi)$ is not the zero polynomial. Hence there exist $0 < p_1, \ldots, p_m < m^{-1}$ for which $Jac(\Psi) \neq 0$. Setting $\mathbf{z} = \Psi(p_1, \ldots, p_m)$, S_a contains a ball about $\mathbf{z}$. ■

Equivalence of Formulations

Theorem 2 $S_a = S_b$.

Proof Fix H, $|V(H)| = i$, and G, $|V(G)| = n$. There are less than $i^2 n^{i-1}$ set mappings $\psi: V(H) \to V(G)$ which are not monomorphisms. Thus

$$A_H(G) - i^2 n^{i-1} \leq B_H(G) \leq A_H(G)$$

and hence

$$a_H(G)[n^i/(n)_i] - i^2 n^{i-1}/(n)_i \leq b_H(G) \leq a_H(G)[n^i/(n)_i].$$

Now $\mathrm{Lim}_n\, n^i/(n)_i = 1$ and $\mathrm{Lim}_n\, i^2 n^{i-1}/(n)_i = 0$. If G_n is a sequence with $|V(G_n)| \to \infty$, then $a_H(G_n) \to \alpha$ iff $b_H(G_n) \to \alpha$. From the definition, $S_a = S_b$. ■

Theorem 3 *S_c is a nonsingular linear transformation of S_b (and hence S_a).*

Proof Let $|V(H)| = i$, $|V(G)| = n$. We claim

$$B_H(G) = \sum_{H_1} B_H(H_1) C_{H_1}(G),$$

where H_1 ranges over all graphs, $|V(H_1)| = |V(H)|$, which contain H as a subgraph. For if $\psi: H \to G$ is a monomorphism $\psi(V(H)) = H_1$, a graph containing H as a subgraph. Conversely, for each H_1 there are $C_{H_1}(G)$ copies of H_1 in G and $B_H(H_1)$ maps ψ into each copy. Dividing by $(n)_i$:

$$b_H(G) = \sum_{H_1} [B_H(H_1)/i!] c_{H_1}(G),$$

giving an explicit transformation from $\mathbf{c}$ to $\mathbf{b}$ for any fixed k. If we order $\{H_1, \ldots, H_m\}$ by number of edges (arbitrarily among graphs with the same number of edges) the coefficient matrix becomes upper triangular with diagonal terms $B_H(H)/i! \neq 0$ and hence the transformation is nonsingular. ■

Topological Properties

Theorem 4 *S_a is arcwise connected (and hence, by Theorems 2 and 3, so are S_b and S_c).*

Proof Let $|V(H_i)| = \alpha_i$ for $1 \le i \le m$. Let $\mathbf{x} = (x_1, \ldots, x_m) \in S_a$. We claim that for $0 \le p \le 1$, $(x_1 p^{\alpha_1}, \ldots, x_m p^{\alpha_m}) \in S_a$. We use that, by Lemma 4

$$a_i(vG + I_w) = a_i(G) r^{\alpha_i} \qquad \text{where} \quad r = v|V(G)|/[v|V(G)| + w].$$

We may find G with $|a_1(G) - x_i|$ arbitrarily small for all i and thence find v, w so that $|p - r|$ is arbitrarily small, thus making $a_i(vG + I_w)$ arbitrarily close to $x_i p^{\alpha_i}$. This completes the claim. The theorem follows as we have given an arc between an arbitrary $\mathbf{x} \in S_a$ and $\mathbf{0}$. ∎

Dependence of Disconnected Graphs

We have shown that the functions a_H, where H runs over connected graphs of size $\le m$ are strongly independent.

Observation Let $G = \sum_{i=1}^{s} G_i$. Then

$$a_G = \prod_{i=1}^{s} a_{G_i}.$$

That is, the functions a_H, H disconnected, are dependent on the a_H, H connected.

Theorem 5 *Suppose $H_1, \ldots, H_n$ represent* all *graphs on $\le k$ vertices, define $\mathbf{a}(G) = (a_1(G), \ldots, a_n(G))$, S_a as before. Then S_a has dimension m equal to the number of connected graphs on $\le k$ vertices.*

Observation Let $H^* = H + I_s$. Then $a_{H^*} = a_H$.

Theorem 6 *Suppose $H_1, \ldots, H_n$ represent all graphs on exactly k vertices, define $\mathbf{a}(G) = (a_1(G), \ldots, a_n(G))$, S_a as before. Then S_a has dimension m equal to the number of connected graphs on $\le k$ vertices.*

The previous theorems also hold for S_b, S_c by equivalence theorems.

Example 2 For $h = 3$, $\dim(S_c) = 3$.

Comments

1. The domain of our functions has been the class $\mathcal{G}$ of all finite graphs. Let $S_a{}^*$ be the set analogous to S_a if we restrict the domain to connected graphs. We claim $S_a{}^* = S_a$ (and similarly $S_b{}^* = S_b$, $S_c{}^* = S_c$).

Clearly $S_a^* \subseteq S_a$. Let $\mathbf{a} \in S_a$ and fix a sequence G_n, $|V(G_n)| \to \infty$, $\mathbf{a}_n = \mathbf{a}(G_n) \to \mathbf{a}$. To each G_n add at most $|V(G_n)| - 1$ edges to form a connected graph G_n^*. One can easily show that, in an asymptotic sense, almost none of the k vertex subgraphs contain any edges of $G_n^* - G_n$ so that $\mathbf{a}(G_n^*) \to \mathbf{a}$.

We may restrict our domain to doubly connected graphs, or similar restrictions, with identical results.

2. It is not known if S_a is locally arcwise connected. In general, the topological nature of S_a is not understood.

3. In the definition of limit points it was required to find a sequence G_n with $|V(G_n)| \to \infty$. It can be shown, using probabilitic methods, that for all sequences G_n, $|V(G_n)| \to \infty$, $\mathbf{a}(G_n) \to \mathbf{a}$, there exists a sequence G_n^* (of which G_n is a subsequence) so that $|V(G_n^*)| = n$ and $\mathbf{a}(G_n^*) \to \mathbf{a}$.

4. The sets S are, in general, not convex. For example—let $k = 3$ and H_i denote the 3-point $(i - 1)$-edge graph, $1 \leq i \leq 4$, then

$$(1, 0, 0, 0), (0, 0, 0, 1) \in S_c \qquad (\text{by } I_s \text{ and } K_s)$$

but

$$(\tfrac{1}{2}, 0, 0, \tfrac{1}{2}) \notin S_c.$$

5. A complete description of convex hull (S) would settle several long standing questions in graph theory. For example, for $k \geq 4$, an ancient conjecture of Erdös is that

$$\min_{\mathbf{c} \in S_c} [c_{I_k} + c_{K_k}] = 2^{1-k}.$$

6. A complete description of S for $k = 3$ appears very difficult.

REFERENCE

1. H. Whitney, The coloring of graphs, *Ann. of Math.* **33** (1932) 688–718.

AMS 05C99, 05A05

Paul Erdös
MATHEMATICS INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST, HUNGARY

László Lovász
BOLYAI INTÉZET
JÓZSEF ATTILA UNIVERSITY
SZEGED, HUNGARY

Joel Spencer
MATHEMATICS DEPARTMENT
STATE UNIVERSITY OF NEW YORK
AT STONY BROOK
STONY BROOK, NEW YORK