Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Daniel Minoli

Combinatorial graph complexity

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **59** (1975), n.6, p. 651–661.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1975_8_59_6_651_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1975.

Teorie combinatorie. — *Combinatorial graph complexity.* Nota di DANIEL MINOLI, presentata (*) dal Socio B. SEGRE.

RIASSUNTO. — Qui si ottiene una misura della complessità di un grafo non orientato; varie misure sono già state proposte, ma esse non soddisfano ad alcune proprietà fondamentali che una siffatta funzione dovrebbe avere, date essenzialmente dal carattere monotonico della complessità rispetto al numero dei vertici, dei lati, e del grado di connessione del grafo. Ecco la nostra definizione: Un cammino tra due vertici v_i and v_j, $v_i \neq v_j$, dicesi proprio se 1) contiene v_i e v_j esattamente una volta, rispettivamente come vertice inziale e finale, e 2) contiene ogni particolare lato al massimo una volta; la complessità $\chi(G)$ di un grafo G viene quindi così definita:

$$\chi(G) = \frac{ne}{n+e} \sum_{(v_i,v_j),\, i<j} \sigma_{ij},$$

dove e = numero dei lati, n = numero dei vertici, σ_{ij} = numero dei cammini propri tra i vertici v_i e v_j. Proprietà di questa complessità vengon qui investigate.

1. INTRODUCTION AND DEFINITIONS

An intrinsic character of a linear graph is its relative 'complexity'. Various measures of complexity have been given, mostly dealing with the information content, and based on certain invariant partitions of the vertex set. See [1], [2], [3] for discussion and further references. However, the measures so far presented fail to satisfy certain fundamental conditions which we would like to have. We construct here a measure which has many appealing properties.

DEFINITION 1. A complexity functions is a positive functional

$$\chi : C \to R$$

[C being the class (category) of graphs].

A complexity function should enjoy the following properties:

1) it should be monotonically increasing on the number of vertices of the graph;

2) it should be monotonically increasing on the number of edges of the graph;

3) it should reflect the degree of connectedness of the graph;

4) it should satisfy our intuitive 'feel' of complexity, assigning a high number to a graph which 'looks' complicated, and viceversa.

These are exactly the conditions that the available measures of complexity fail to satisfy.

(*) Nella seduta del 15 novembre 1975.

Our paper is anti-climactic, establishing the most important result with Definition 4 below and then proceeding to show how the characterizing properties presented above are satisfied.

V and E denote, throughout, the vertex and the edge set respectively. In the present work, complexity will be defined only for normal graphs, but extensions are possible.

DEFINITION 2. A connected, undirected graph with no multiple edges and no self loops, and with V having more than two points, is called a *normal graph.*

DEFINITION 3. A path between two vertices v_i , v_j (with $v_i \neq v_j$) is proper if

a) it contains v_i and v_j exactly once, and as initial and final vertex respectively;

b) it contains any particular edge no more than once.

Let $n = |V|$, V the vertex set; $e = |E|$, E the edge set; $\Delta = \{(v_i , v_i) , v_i \in V\}$.

DEFINITION 4. Let $G = (V , E)$ be a normal graph. We define

$$\chi(G) = \frac{ne}{n+e} \sum_{\substack{(v_i, v_j) \\ i>j}} \sum_{k=1}^{e} \gamma_k(i , j) ,$$

with $\gamma_k(i , j)$ = number of proper paths of length k from v_i to v_j, and the outer summation taken over all $(v_i , v_j) \in (V \times V) - \Delta$, to be the *combinational complexity function*, or, in short, the *complexity function.*

Naturally we could lump together all proper paths between a pair of vertices (v_i , v_j) as σ_{ij} and write

$$\chi(G) = \frac{ne}{n+e} \sum_{\substack{(v_i, v_j) \\ i>j}} \sigma_{ij}$$

but the former approach is easier to work with to obtain closed form expressions for particular graphs.

From now on, unless there is danger of confusion, we shall abbreviate $\gamma_k(i , j)$ by γ_k, and whenever we say 'path' we mean 'proper path'. If a graph G_1 is isomorphic with a graph G_2 we denote this as $G_1 \simeq G_2$; if they are not isomorphic, we denote this as $G_1 \not\simeq G_2$. By 'graph' we always mean a *normal graph.*

THEOREM 1. $G_1 \simeq G_2$ *implies that* $\chi(G_1) = \chi(G_2)$.

Proof. Since $e_1 = e_2 = e$ and $n_1 = n_2 = n$, we need only study

$$\Gamma_{G_1} = \sum_{\substack{(v_i, v_j) \\ j>i}} \sigma_{ij}^{G_1} \quad , \qquad \sigma_{ij}^{G_1} = \sum_{k=1}^{e} \gamma_k^{G_1}(i , j) \quad ,$$

$$\Gamma_{G_2} = \sum_{\substack{(v_p, v_g) \\ g>p}} \sigma_{pg}^{G_2} \quad , \qquad \sigma_{pg}^{G_2} = \sum_{k=1}^{e} \gamma_k^{G_2}(p , g) \quad .$$

To every term of the form (v_{i_0}, v_{j_0}) in the sum for Γ_{G_1} there corresponds a term involving (s_{p_0}, s_{g_0}) in the sum for Γ_{G_2}, with s_{p_0} the image of v_{i_0} under the isomorphism, and s_{g_0} the image of v_{j_0}; (if $p_0 > q_0$ this term does not actually appear in the second sum but (s_{g_0}, s_{p_0}) appears and $\sigma_{p_0 g_0} = \sigma_{g_0 p_0}$; without loss of generality we take $p_0 < q_0$). Similarly, for every term involving (s_{p_0}, s_{g_0}) in Γ_{G_1} there corresponds a term for (v_{i_0}, v_{j_0}) in Γ_{G_1}.

The discussion above implies that working with reference to Γ_{G_1} it suffices to show

$$\sigma_{i_0 j_0}^{G_1} = \sigma_{p_0 g_0}^{G_2}$$

for all (i_0, j_0). Assume that

$$\sigma_{i_0 j_0}^{G_1} \neq \sigma_{p_0 g_0}^{G_2}$$

this means
$$\sum_{k=1}^{e} \gamma_k (i_0, j_0) \neq \sum_{k=1}^{e} \partial_k (p_0, g_0)$$

or, in other words, there is a k such that $\gamma_k \neq \partial_k$. Without loss of generality assume

$$(*) \qquad \gamma_k > \partial_k .$$

Now, say $v_{i_0} = v_0, v_1, v_2, \cdots, v_t = v_{j_0}$ determines a proper path in G_1; there are $\gamma_k (i_0, j_0)$ such paths G_1. But then $s_{p_0} = s_0, s_1, s_2, \cdots, s_t = s_{g_0}$, s_w being the image of v_w, also determines a proper path in G_2, since isomorphism preserves incidence at all vertices. Consequently,

$$\partial_k \geq \gamma_k ,$$

but this is a contradiction with assumption (*); hence,

$$\sigma_{i_0 j_0}^{G_1} = \sigma_{p_0 g_0}^{G_2} .$$

This is true for all (i_0, j_0), which establishes the theorem. The converse is not true, as we show later.

THEOREM 2. *The fundamental theorem on graph complexity. Let* G *be a graph on n vertices and e edges,* H *a subgraph on n' vertices and e' edges. If the complexity of* H *is* $\chi(H)$, *then*

$$\chi(G) \geq \frac{ne}{n+e} \left[\frac{n'+e'}{n' e'} \chi(H) + (e - e') \right] .$$

Proof. Since H has complexity $\chi(H)$ it contains, by definition of complexity

$$\frac{n'+e'}{n'e'} \chi(H)$$

proper paths. The remaining edges, $e - e'$, provide *at least* $e - e'$ other paths.

2. Properties of the complexity function

We now want to see if and how the four properties listed above for the defined complexity function hold. We can settle property three at once, since the definition obviously takes into account connectedness of the graph. We investigate now properties one and two.

First we analyse $$\mathrm{Y}(n, e) = \frac{ne}{n+e}.$$

Consider $\mathrm{Y}(n, e)$ as a continuous function in $\mathrm{R}^2 - \{(0, 0)\}$.

$$\frac{\delta}{\delta n}\frac{ne}{n+e} = \frac{(n+e)e - ne}{(n+e)^2} = \frac{e^2}{(n+e)^2} > 0 \quad \text{for all } e.$$

Consequently $\mathrm{Y}(n, e)$ is increasing for all fixed e_0:

$$\frac{\delta}{\delta n}\frac{ne}{n+e} = \frac{(n+e)n - ne}{(n+e)^2} = \frac{n^2}{(n+e)^2} > 0 \quad \text{for all } n.$$

Consequently $\mathrm{Y}(n, e)$ is increasing for all fixed n_0. Finally, for $n_1 \leq n_2$, $e_1 \leq e_2$.

$$\frac{n_2 e_2}{n_2 + e_2} \geq \frac{n_2 e_1}{n_2 + e_1} \geq \frac{n_1 e_1}{n_1 + e_1}$$

with strict inequality if at least one of the above conditions on n or e is a strict inequality.

THEOREM 3. *Let* G *be a given graph with n vertices and e edges. Let* G′ *be a graph obtained from* G *by adding edges. Then*

$$\chi(\mathrm{G}') \geq \chi(\mathrm{G}).$$

Proof. Since e increases and n remains constant by construction $\mathrm{Y}(n, e)$ increases. $\Gamma_{\mathrm{G}'}$ exceeds Γ_{G} since

a) we are taking the outer sum over the same set;

b) we have added edges hence, since G′ is obtained from G, in addition to the proper paths contained in G, we must have some new proper paths. This proves the statement.

COROLLARY 1. *Let* G *be a graph with n vertices and e edges. Let* G′ *be obtained from* G *by adding vertices such that the resultant graph is connected* (*we have not defined complexity for disconnected graphs*). *Then*

$$\chi(\mathrm{G}') \geq \chi(\mathrm{G}).$$

Proof. n has increased; since we require connectedness, e must also have increased. Invoke Theorem 4, whence the result.

Let t_n and s_n be two non-isomorphic trees on n vertices; we show later that $\chi(t_n) = \chi(s_n)$. We can then state

THEOREM 4. *K_n is the most complex graph on n vertices; t_n, a tree on n vertices, is the least complex graph on n vertices.*

Proof. K_n must necessarily have the largest number of proper paths, since every possible edge is present. On the other hand, t_n has the least number of paths. Finally

$$e_{K_n} = a \geq e_{t_n} = b$$

implies that $Y(n, a) \geq Y(n, b)$. Putting all of this together we get the result.

We may now want to know what happens as we go from t_0 to K_n. For a detailed analysis of this problem, we refer the reader to [4]; the final result is as follows:

As a result of Theorem 3 we know that if we *take a spanning subtree on n vertices and we successively add edges we increase the complexity* (hence Theorem 4), but we cannot, in general, establish the relationship between the complexity of a graph obtained by *adding e_1 edges to a spanning subtree and the complexity of a graph obtained by adding e_2 edges, $e_1 < e_2$ to a different spanning subtree on n vertices.*

3. CLOSED FORM FORMULAE

We now present exact formulae for the complexity of well known graphs.

THEOREM 5. *Let G be a tree on n vertices. Then*

$$\chi(G) = \frac{n^2 (n-1)^2}{4n-2} .$$

Proof. Without loss of generality we work with a string.

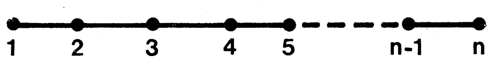

Fig. 1.

Vertices : n. — Edges : $n - 1$. — Proper paths: as for the following Table I.

TABLE I.

	1	2	3	$4 \cdots n-1 \leftarrow k$	
12	1	0	0	$0 \cdots 0$	
13	0	1	0	$0 \cdots 0$	
14	0	0	1	$0 \cdots 0$	
.	.				
.	.				
23	1	0	0	$0 \cdots 0$	$\leftarrow \Upsilon_k$
.	.				
.	.				
$n-1, n$					

Every row contributes one path; thre are $\frac{1}{2} n (n - 1)$ rows, hence

$$\chi(G) = \frac{ne}{n+e} \Sigma\Sigma\gamma_k = \frac{n(n-1)}{n+n-1} \frac{(n-1)n}{2} .$$

With this theorem we can prove the assertion we have made before that the converse of Theorem 1 is false: simply take two non-isomorphic trees on n vertices. In a similar way we prove

THEOREM 6. *If* G *is a cycle on* n *vertices then*

$$\chi(G) = \frac{n^2(n-1)}{2} .$$

THEOREM 7. *For the complete graph on* n *vertices,*

$$\chi(K_n) = \frac{1}{2} \frac{n^4 - 2n^3 + n^2}{n+1} (n-2)! \sum_{r=0}^{n-2} \frac{1}{(n - r \cdot 2)!} .$$

Proof. Vertices: n. — Edges: $\frac{1}{2}(n-1)$. — Proper paths: as in Table II.

TABLE II.

	1	2	3	4	5 ···	0	$\leftarrow k$
12	1	$\frac{(n-2)!}{(n-2-1)!}$	$\frac{(n-2)!}{(n-2-2)!}$	$\frac{(n-2)!}{(n-2-3)!}$	$\frac{(n-2)!}{(n-2-4)!}$	$\frac{(n-2)!}{0!}$	
13	1	$\frac{(n-2)!}{(n-2-1)!}$	$\frac{(n-2)!}{(n-2-2)!}$	—	—	—	
14	1	$\frac{(n-2)!}{(n-2-1)!}$	—	—	—	—	
⋮							
23	1	—	—	—	—	—	$\leftarrow \gamma_k$
⋮							
$n-1, n$	1	—	—	—	—	—	

Obviously every row is identical since we have symmetry, hence we consider only paths between the pair (1, 2). The first entry in this row is clearly 1. To determine the second entry we must determine how many paths of length two there are between vertex 1 and vertex 2. Let

$$x \to z \to t \to \cdots \to v \to y$$

stand for the path from vertex x to vertex y through the vertices $z, t, \cdots, v$, in the given order. Then, in the case above we would have

$$1 \to 3 \to 2 , \quad 1 \to 4 \to 2 , \quad 1 \to 5 \to 2 , \cdots , 1 \to n \to 2 .$$

This is equivalent to determine in how many ways we can pick one symbol —the middle symbol in the above formal description of a path—out of $n-2$ (we have to exclude 1 and 2); this is

$$\gamma_2 = \frac{(n-2)!}{(n-2-1)!} .$$

Similarly, for paths of length three we have

$$1 \to 4 \to 5 \to 2 , \quad 1 \to 5 \to 4 \to 2 , \quad 1 \to 3 \to 4 \to 2 , \quad 1 \to 4 \to 3 \to 2 ,$$
$$1 \to 3 \to 5 \to 2 , \quad 1 \to 5 \to 3 \to 2 , \cdots .$$

That is, how to take two symbols out of $n-2$, or

$$\gamma_3 = \frac{(n-2)!}{(n-2-2)!} .$$

In general, we get
$$\gamma_{i+1} = \frac{(n-2-i)!}{(n-2)!} .$$

The total number of paths contributed by *each row* in Table II is

$$\sum_{r=0}^{n-2} \frac{(n-2-r)!}{(n-2)!} .$$

There are $\frac{1}{2} n (n-1)/2$ rows. It follows that

$$\chi (G) = \frac{n \frac{n(n-1)}{2}}{\frac{2n + n(n-1)}{2}} \frac{(n-1) n}{2} (n-2)! \sum_{r=0}^{n-2} \frac{1}{(n-2-r)!} ,$$

from which the result can be obtained by algebraic simplification.

Note that for $n = 2$ the formulas for t_2 and K_2 coincide; similarly for c_3 and K_3. The following theorem may be established:

THEOREM 8. *For a bipartite graph* $K_{m,m}$,

$$\chi (K_{m,m}) = \frac{2m^2}{2+m} \left[m^2 + (m!)^2 \left\{ \sum_{j=0}^{m-2} \frac{1}{[(m-2-j)!]^2} \left[1 + \frac{1}{m-j-1} \right] \right\} \right] .$$

4. UPPER AND LOWER BOUNDS

The complexity of a graph G admits in a natural way lower and upper bounds. Some of these bounds will be very insensitive, but they are useful either to prove other theorems, or to give an absolute bound just in term of very little information on G, say n. The next theorem is of this type.

THEOREM 9. *For any normal graph* G *on* n *vertices*

$$\chi (G) \geq \frac{n^3}{2(n^2+n)} \qquad (B\ 1).$$

Proof. A normal graph is connected, hence $2e \geq n$. Since a normal graph has no multiple edges

$$e \leq \frac{n(n-1)}{2} .$$

It follows that

$$\chi(G) = \frac{ne}{n+e} (\Sigma\Sigma\, \gamma_k) \geq \frac{n(n/2)}{n+(n^2-n)/2} (\Sigma\Sigma\, \gamma_k) .$$

But the sum of all proper paths is *at least the number of edges* e (an abundant oversimplification). Finally

$$\chi(G) \geq \frac{n^2}{n^2+n}\, e \geq \frac{n^2}{n^2+n} \cdot \frac{n}{2} = \frac{n^3}{2(n^2+n)} .$$

COROLLARY 2.

$$\chi(G) \geq \frac{n^2}{n^2+n}\, e \qquad (B\,2) .$$

While B 1 gives a lower bound for $\chi(G)$ by knowing n only, B2 also utilizes the information contained in e; hence is a better lower bound. Note that for large n, B1 $\sim n/2$.

THEOREM 10. *For any normal graph* G *on* n *vertices*

$$\chi(G) \geq \frac{2(n-1)^2}{n+1} \qquad (B\,3) .$$

Proof. Since G is connected,

$$e \geq n-1 \quad , \quad e \leq \frac{n^2-n}{2} ,$$

$$\chi(G) = \frac{ne}{n+e} (\Sigma\Sigma\, \gamma_k) \geq \frac{n(n-1)}{n+(n^2-n)/2} (\Sigma\Sigma\, \gamma_k) =$$

$$= \frac{2n(n-1)}{n(n+1)} (\Sigma\Sigma\, \gamma_k) \geq \frac{2(n-1)}{(n+1)}\, e \geq \frac{2(n-1)^2}{(n+1)} .$$

Note that for large n, B 3 $\sim 2(n-1)$, which is better than the asymptotic behavior of B 1.

COROLLARY 3.

$$\chi(G) \geq \frac{(n+1)}{2(n-1)}\, e \qquad (B\,4) .$$

B 1, B 2, B 3, B 4 are attained for K_2.

THEOREM 11.

$$\chi(G) \geq \frac{n+e}{ne} \{2e-n+1\} \qquad (B\,5).$$

Proof. Let $c(G)$ be the cyclomatic number; every edge e provides a proper path, and every independent cycle provides *at least another path*, which differs from the one above. At this point we can only say ' at least another path ' since the other ' half ' cycle, which we would like to count, may already have been counted as a single edge. For exemple, in fig. 2, the cycle only provides one new path between vertex 1 and 2 (it provides two new paths between vertex 1 and 3). Hence

$$\chi(G) \geq \frac{ne}{n+e}\{e + c(G)\} = \frac{ne}{n+e}\{e + (e - n + 1)\} \frac{ne}{n+e} = (2e - n + 1).$$

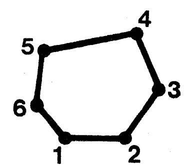

Fig. 2

B5 is attained for K_2. — We now derive a better bound.

THEOREM 12. *If* $e \geq n$, *then*

$$\chi(G) \geq \frac{ne}{n+e}\left\{\frac{(n-1)n}{2} + (e - n + 1)\right\} \qquad (B\,6).$$

Proof: The graph in question is a supergraph of a spanning tree on n vertices; a subtree of this kind provides $(n-1)\,n/2$ proper paths, as we saw in Section 3; but there are $(e - n + 1)$ other edges providing proper paths. Using the principle of the Fundamental Theorem, we obtain the desired result.

The condition $e \geq n$ is only required to insure that the second term is reasonably large, otherwise the bound is poor.

The following bounds may also be established (see [4]). Let φ be an integer. Define

$$\delta_\varphi = 0 \text{ if } \varphi = 0, \qquad \delta_\varphi = 1 \text{ if } \varphi \geq 1.$$

Let $\varphi = 3n - 2e - 3$ be larger than zero, and let

$$\tilde{w} = \min \sum_{\varphi_i \in \mathbf{P}} \left[\frac{(3 + \varphi_i - 1)(3 + \varphi_i)}{2} - 1\right] \delta_{\varphi_i}$$

where the minimum is taken over all partitions $\mathbf{P} = \{\varphi_1, \varphi_2, \cdots, \varphi_t\}$ of φ into *at most* $c(G)$ summands. Then

$$(B\,7) \qquad \chi(G) \geq \frac{ne}{n+e}\left[\frac{(n-1)n}{2} + (e - n + 1) + \tilde{w}\right],$$

$$(B\,8) \qquad \chi(G) \geq \frac{ne}{n+e}\, G(e - n + 1),$$

$$(B\,9) \qquad \chi(G) \geq \frac{ne}{n+e}\left[2z^* + 5\sum_{j=1}^{z^*-1} 2^{j+1}(z^* - j)\right],$$

where $z^* = \min(c, y_n^*)$ whith c the cyclomatic number and

$$y_n^* = \begin{cases} \left[\frac{n}{2}\right] \text{ if } n \text{ odd,} \\ \left[\frac{n}{2}\right] - 1 \text{ if } n \text{ even.} \end{cases}$$

We compare these bounds for a particular example, along with the actual complexity.

EXAMPLE:

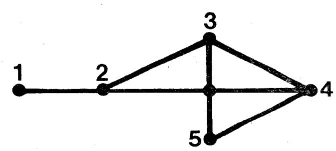

Fig. 3.

Pair (v_i, v_j), $j > i$	Number of proper paths $= \sigma_{ij}$
12	1
13	3
14	3
15	3
23	3
24	3
25	3
34	3
35	3
45	3

$$\chi(G) = \frac{ne}{n+e} \sum_{(v_i, v_j)} \sigma_{ij} = \frac{5 \cdot 6}{5+6} \sum_{(v_i, v_j)} \sigma_{ij} = \frac{30}{11}(28) = 76.3$$

Bound Bx	Value
B 1	2.08
B 2	5.00
B 3	5.33
B 4	8.00
B 5	21.80
B 6	30.00
B 7	Does not apply: $\varphi = 0$
B 8	32.70
B 9	65.45

Notes: $n = 5$; $e = 6$; $c(G) = 2$; $y_n^* = 2$; $z^* = 2$. — The situation with upper bounds is much simpler.

THEOREM 11. *Let* G *be a normal graph on n vertices. Then*

$$\chi(\mathrm{G}) \leq \frac{1}{2} \frac{n^4 - 2n^3 + n^2}{n+1} (n-2)! \sum_{r=0}^{n-2} \frac{1}{(n-2-r)!} = f.$$

Proof. $\chi(\mathrm{G}) \leq \chi(\mathrm{K}_n)$, and $\chi(\mathrm{K}_n) = f$.

COROLLARY 4. *For a normal graph* G

$$\chi(\mathrm{G}) < \frac{1}{2} \frac{n^4 - 2n^3 + n^2}{n+1} (n-2)! \, (e_{\mathrm{NEP}}) .$$

Proof.

$$\sum_{r=0}^{n-2} \frac{1}{(n-2-r)!} = \frac{1}{(n-2)!} + \frac{1}{(n-3)!} + \cdots + \frac{1}{2!} + \frac{1}{1!} + \frac{1}{0!} < e_{\mathrm{NEP}} = 2.718 \cdots$$

CONCLUSION

The objective of producing a measure of complexity satisfying the requirements outlined above has been accomplished. This complexity enjoys other useful properties not discussed here. The definition can be extended to non-normal, non-oriented graphs, and, also, to oriented graphs; such a complexity on oriented graphs can then be used to define an appealing measure for entropy.

REFERENCES

[1] C. W. MARSHALL (1971) – *Applied Graph Theory*, Wiley–Interscience, New York.

[2] A. MOWSHOWITZ (1968) – *Entropy and the Complexity of Graphs*. II, « Bull. Math. Biophys. », *30*, 225–240.

[3] E. TRUCCO (1956) – *A Note on the Information Content of Graphs*, « Bull. Math. Biophys. », 129–135.

[4] D. MINOLI – *Graph Complexity*, Unpublished Thesis.