Refinement of Metrics: Erdős Number, a Case Study

K. Lock, W.Y. Pong, and A. Wittmond*

Abstract - We introduce a concept called refinement and develop two different ways of refining metrics. By applying these methods we produce several refinements of the shortest-path distance on the collaboration graph and hence a couple of new versions of the Erdős number.

Keywords : refinement; metrics; Erdős number; monoid norms; monotonic monoid norms

Mathematics Subject Classification (2010) **:** 05C12; 05C90

1 Introduction

Our investigation was motivated by a simple goal: to find a "better" version of the Erdős number.

The **Erdős number** of a person can be defined recursively as follows: Paul Erdős has Erdős number 0. A person other than Erdős himself has Erdős number one more than the smallest Erdős number among his/her coauthors. If none of the person's coauthors have an Erdős number, then neither does that person. Equivalently, the Erdős number of a person is the shortest-path distance between that person and Paul Erdős in the graph where there is an edge between two people if they are among the authors of a paper in mathematics. We refer to this graph as the **collaboration graph**. The American Mathematical Society provides an online tool[1] for computing the shortest-path distance between any two mathematicians in the collaboration graph. Oakland University hosts the Erdős number project[2] which provides many interesting facts and data about the Erdős number.

The shortest-path distance is a measure of closeness between nodes in a graph. One may argue, however, that it is an inadequate measure of closeness between collaborators. For instance, it is reasonable to say that the more joint articles between the two people, the closer they are as collaborators. However, such a natural idea is completely ignored by the shortest-path distance. The ratio of the number of joint articles to the total number of publications between two authors is another piece of information that can be used in measuring distances of collaborators. These considerations suggest one to view

*This work was supported by a PUMP research grant (NSF grant No. DMS-1247679). The first and the third author were undergraduate students at California State University, Dominguez Hills.

[1] `https://mathscinet.ams.org/mathscinet/freeTools.html?version=2`

[2] `https://oakland.edu/enp/`

the collaboration graph as a weighted graph rather than just a simple graph. In a finite weighted graph, the lightest-path distance between two nodes in the same connected component is the minimum path weight between the two nodes. The resistance distance is another metric on weighted graphs in which every path between two nodes contributes some decrement of the distance between them. The idea is that the more paths there are connecting the two nodes, the less “resistance” there is to travel from one to the other and hence the closer they are. Finding effective resistance between nodes in an electric circuit is certainly familiar to engineers, while viewing it as a metric on graphs is no strange business to graph theorists either. We refer the reader to the book [4] by Bollobás and the article [9] by Shapiro for more information. More recently, Chebotarev constructed a family of graph-geodetic distances [5] in which the lightest-path distance and the resistance distance correspond to the two extreme cases of the parameter. Using the resistance distance to measure closeness of collaborators was considered in [1], [2] and [10]. Some non-metrical generalizations of the Erdős number that measure proximity between nodes in weighted networks was proposed and studied in [8]. Since it is of secondary interest to us in this article, we mention in passing that algorithms for computing the lightest-path distance are well-known [6] and those for computing resistance distance have also been widely studied. The article [11] contains more references on this topics.

There is one aspect of using either the lightest-path distance or the resistance distance to define Erdős numbers that is unsatisfactory to us, namely, the relative closeness between authors given by these metrics may contradict the one given by the shortest-path distance: A may be closer to B than to C according the shortest-path distance but the exact opposite may be true for either of the two distances aforementioned. Because of this, we set our goal to finding metrics on the collaboration graph that, in some fashion, take into account the number of joint articles but do not contradict the shortest path distance closeness.

The rest of this article is organized as follows: in Section 2 we introduce a notion called monoid norm and use it to unify various constructions of metrics. In Section 3 we introduce the refinement relation on functions defined on a Cartesian product with codomain a totally ordered set. We then show how to produce refinements of a metric by another metric. We also identify a condition under which the refining process can be iterated. Section 4 is devoted to a particular kind of refinement of the shortest-path distance. Unlike the constructions given in Section 3, the additional functions used in the refining process are no longer metrics. However, it is crucial that the metric being refined is the shortest-path distance. Lastly, in Section 5, we compute the new Erdős numbers of a few mathematicians corresponding to different refinements of the shortest path distance. We end the article by proposing another edge weight function which seems to be appropriate for the purpose of refining the Erdős number.

2 Monoid norms

To produce metrics that fit our requirements set forth in the introduction, we use several basic constructions of metrics. What seems to be new to us here is the realization that all these constructions can be unified into a single one. This led us to the following pair of notions. Let $(M, +, 0_M)$ be a monoid (written additively). We call a function μ from M to $\mathbb{R}$ a **monoid norm** if

1. $\mu(x) \geqslant 0$ for any $x \in M$;
2. $\mu(x) = 0$ if and only if $x = 0_M$; and
3. $\mu(x + y) \leqslant \mu(x) + \mu(y)$ for any $x, y \in M$. (subadditivity)

A partially ordered monoid is a monoid $(M, +, 0_M)$ equipped with a partial order $\leqslant_M$ on M that respects translation, i.e. $x \leqslant_M y$ implies $x + z \leqslant_M y + z$ for any $x, y, z \in M$. As an example, let $\mathbb{R}^n_{\geqslant 0}$ $(n \geqslant 1)$ be the set of n-tuples of non-negative real numbers. For $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}' = (x'_1, \ldots, x'_n)$ in $\mathbb{R}^n_{\geqslant 0}$, let $\mathbf{x} \leqslant \mathbf{x}'$ if $x_i \leqslant x'_i$ in the usual order of real numbers for each $1 \leqslant i \leqslant n$. The relation $\leqslant$ thus defined is called the product order on $\mathbb{R}^n_{\geqslant 0}$. It is straightforward to check that $\mathbb{R}^n_{\geqslant 0}$ equipped with component-wise addition and the product order is a partially ordered monoid. We call a monoid norm μ on a partially ordered monoid M **monotonic** if $x \leqslant_M y$ implies $\mu(x) \leqslant \mu(y)$. The reader will likely recognize that the names of these notions are taken from their counterparts for real vector spaces. The only difference is that the homogeneity property of norms, that is $\mu(\alpha x) = |\alpha|\mu(x)$ for $\alpha \in \mathbb{R}$, which does not make sense for monoids in general, is being dropped.

Example 2.1 The following functions are monotonic monoid norms (the first three functions are defined for $\mathbb{R}_{\geqslant 0}$ and the last one is defined for $\mathbb{R}^n_{\geqslant 0}$ $(n \geqslant 1)$):

(i) $\mu(x) = \alpha x$ $(\alpha > 0)$.

(ii) $\mu(x) = \lceil x \rceil$ where $\lceil x \rceil$ denotes the least integer no smaller than x.

(iii) $\mu(x) = \dfrac{x}{1 + x}$.

(iv) $\mu(x_1, \ldots, x_n) = x_1 + \cdots + x_n$.

We will verify the last function is a monotonic monoid norm and leave the verification of the other three to the reader. For any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n_{\geqslant 0}$, since each x_i is non-negative, $\mu(\mathbf{x}) = \sum_{i=1}^n x_i \geqslant 0$ and $\mu(\mathbf{x}) = 0$ if and only if each $x_i = 0$, i.e. $\mathbf{x} = \mathbf{0}$. Moreover, since $\mathbf{x} + \mathbf{x}' = (x_1 + x'_1, \ldots, x_n + x'_n)$, so $\mu(\mathbf{x} + \mathbf{x}')$ actually equals $\mu(\mathbf{x}) + \mu(\mathbf{x}')$. Thus, μ is a monoid norm on $\mathbb{R}^n_{\geqslant 0}$. Finally, suppose $\mathbf{x} \leqslant \mathbf{x}'$; that is, $x_i \leqslant x'_i$ for each $1 \leqslant i \leqslant n$ and so $\mu(\mathbf{x}) = \sum_{i=1}^n x_i \leqslant \sum_{i=1}^n x'_i = \mu(\mathbf{x}')$. Therefore, μ is monotonic.

Some monoid norms on $\mathbb{R}^n_{\geqslant 0}$ are clearly not restrictions of norms on $\mathbb{R}^n$. One example is the ceiling function $x \mapsto \lceil x \rceil$. Another example is the function defined by $\mu(\mathbf{0}) = 0$ and $\mu(\mathbf{x}) = 1$ for all $\mathbf{x} \neq \mathbf{0}$. Both of them fail the homogeneity property for being a norm. On the other hand, many familiar norms on $\mathbb{R}^n$, e.g. the ℓ_p norms, are monotonic on $\mathbb{R}^n_{\geqslant 0}$. In the literature, norms that are monotonic on various orthants are studied under the name of orthant-monotonic norms. Despite the fact that there are numerous characterizations of these norms [3, 7], we were unable to find in the literature an explicitly given norm on $\mathbb{R}^n$ that is not monotonic on $\mathbb{R}^n_{\geqslant 0}$. So it may be worthwhile to include a family of examples here.

Example 2.2 The function $\nu_n \colon \mathbb{R}^n \to \mathbb{R}$ $(n \geqslant 2)$ defined by

$$\nu_n(\mathbf{x}) = \max_{1 \leqslant k \leqslant n} |x_k| + \sum_{i<j} |x_i - x_j|$$

is a norm on $\mathbb{R}^n$ restricting to a monoid norm on $\mathbb{R}^n_{\geqslant 0}$ that is not monotonic. For instance, $(0, 2, \dots, 2) \leqslant (1, 2, \dots, 2)$ in $\mathbb{R}^n_{\geqslant 0}$ and yet

$$\begin{aligned}\nu_n(0, 2, \dots, 2) &= 2 + 2(n-1) > 2 + (n-1) \\ &= \nu_n(1, 2, \dots, 2).\end{aligned}$$

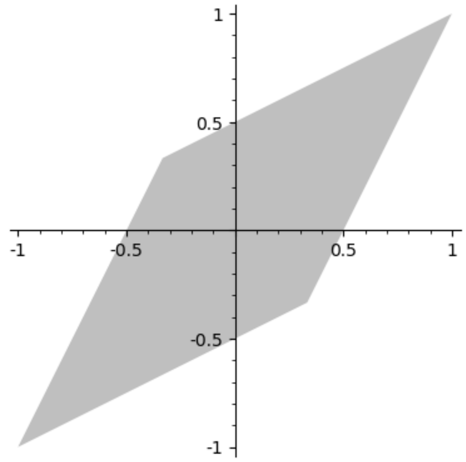


Figure 1: The unit ball of ν_2.

The following statement is the key in unifying various constructions of metrics by the notion of monotonic monoid norms.

Theorem 2.3 *Let (E_i, d_i), $1 \leqslant i \leqslant n$, be metric spaces and μ be a monotonic monoid norm on $\mathbb{R}^n_{\geqslant 0}$. Then $\mu \circ \mathbf{d}$, where $\mathbf{d} = \prod_{i=1}^n d_i$, is a metric on the Cartesian product $\prod_{i=1}^n E_i$.*

Proof. Since each d_i is non-negative, the range of $\mathbf{d}$ is in $\mathbb{R}^n_{\geqslant 0}$. Thus, the function $\mu \circ \mathbf{d}$ is well defined. Since μ takes only non-negative values, so does $\mu \circ \mathbf{d}$ and since each d_i is symmetric, $\mathbf{d}(\mathbf{p}, \mathbf{q}) = \mathbf{d}(\mathbf{q}, \mathbf{p})$ for any $\mathbf{p}, \mathbf{q} \in \prod_{i=1}^n E_i$.

For $1 \leqslant i \leqslant n$, since $d_i(p_i, p_i) = 0$ for every $p_i \in E_i$, $\mathbf{d}(\mathbf{p}, \mathbf{p}) = \mathbf{0}$ and hence $\mu \circ \mathbf{d}(\mathbf{p}, \mathbf{p}) = 0$ for each $\mathbf{p} \in \prod_{i=1}^n E_i$. Because μ only maps $\mathbf{0}$ to 0, $\mu \circ \mathbf{d}(\mathbf{p}, \mathbf{q}) = 0$ implies $\mathbf{d}(\mathbf{p}, \mathbf{q}) = \mathbf{0}$. The last equation means $d_i(p_i, q_i) = 0$ $(1 \leqslant i \leqslant n)$ and so $p_i = q_i$ since each d_i is a metric. Therefore, $\mathbf{p} = \mathbf{q}$.

Finally, since each d_i satisfies the triangle inequality, $\mathbf{d}(\mathbf{p}, \mathbf{q}) \leqslant \mathbf{d}(\mathbf{p}, \mathbf{r}) + \mathbf{d}(\mathbf{r}, \mathbf{q})$ in the product order of $\mathbb{R}^n_{\geqslant 0}$. It then follows from the monotonicity and the subadditivity of μ that

$$\mu \circ \mathbf{d}(\mathbf{p}, \mathbf{q}) \leqslant \mu\left(\mathbf{d}(\mathbf{p}, \mathbf{r}) + \mathbf{d}(\mathbf{r}, \mathbf{q})\right) \leqslant \mu \circ \mathbf{d}(\mathbf{p}, \mathbf{r}) + \mu \circ \mathbf{d}(\mathbf{r}, \mathbf{q}).$$

Thus, $\mu \circ \mathbf{d}$ is indeed a metric on $\prod_{i=1}^n E_i$. □

The next few propositions about metrics, in the light of Theorem 2.3, are all consequences of the fact that the functions in Example 2.1 are monotonic monoid norms.

Proposition 2.4 *If d is a metric, then so are αd $(\alpha > 0)$, $\lceil d \rceil$ and $d_\flat := \dfrac{d}{1+d}$.*

Proposition 2.5 *Let (E_i, d_i), $1 \leqslant i \leqslant n$, be metric spaces, then the map d defined by $d(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} d_i(p_i, q_i)$ is a metric on $\prod_{i=1}^{n} E_i$.*

Corollary 2.6 *The sum of finitely many metrics on a set E is a metric on E.*

Proof. Let $d_1, \dots, d_n$ be metrics on E. By Proposition 2.5, $d(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} d_i(p_i, q_i)$ is a metric on E^n and hence on its diagonal which can be identified with E itself via $p \mapsto \iota(p) = (p, \dots, p)$. Thus, $d(p, q) = d(\iota(p), \iota(q)) = \sum_{i=1}^{n} d_i(p, q)$ is a metric on E. $\square$

3 Refinements

We propose the following notion for functions from a Cartesian product to a totally ordered set.

Definition 3.1 *Let f and f' be functions from a Cartesian product X to a totally ordered set $(T, \leqslant)$. We say that f'* **refines** *f if $f'(\mathbf{x}) < f'(\mathbf{x}')$ whenever $f(\mathbf{x}) < f(\mathbf{x}')$ for any $\mathbf{x}, \mathbf{x}' \in X$ with all but one coordinate the same. We write $f' \preceq f$ if f' refines f.*

The relation $\preceq$ is clearly reflexive and transitive. It is however not anti-symmetric. For example, the identity function of $\mathbb{R}$ and twice this function refine each other. More generally, if σ is an order-preserving embedding from the range of f to T then f refines $\sigma \circ f$ and vice versa. By a **refinement class**, we mean an equivalence class of the equivalence relation in which two functions are equivalent if they refine each other. Note that the refinement relation $\preceq$ induces a partial order, still denoted by $\preceq$, on refinement classes.

For unary functions, f' refines f means f' never contradicts f on strict inequalities. For binary functions, f' refines f if and only if for all $(x_1, x_2) \in X$, $f'(x_1, \_\_)$ refines $f(x_1, \_\_)$ and $f'(\_\_, x_2)$ refines $f(\_\_, x_2)$. Metrics are symmetric binary functions, so for metrics d and d' on the same set E, d' refines d means for any $p, q, r \in E$, $d'(p, q) < d'(p, r)$ whenever $d(p, q) < d(p, r)$ or, equivalently, $d'(p, \_\_)$ refines $d(p, \_\_)$ for each $p \in E$.

There is also a graph theoretic interpretation of refinement. Suppose X is the Cartesian product of a family $\{X_i\}_{i \in I}$ of sets. Let K_i be the complete graph with vertex set X_i ($i \in I$). Let G_X be the Cartesian product of the family $\{K_i\}_{i \in I}$ of graphs. In other words, G_X is the graph with vertex set X where two elements of X are adjacent in G_X if and only if they differ at exactly one $i \in I$. Note that G_X is connected if the index set I is a finite set. On the other hand, G_X is disconnected if each X_i is nonempty and $|X_i| \geqslant 2$ for infinitely many $i \in I$. A function f from X to a totally ordered set T can be viewed as a "potential function" on G_X that gives a directed graph structure on G_X: an edge of G_X between $\mathbf{x}$ and $\mathbf{x}'$ becomes an arc from $\mathbf{x}$ to $\mathbf{x}'$ if $f(\mathbf{x}') \leqslant f(\mathbf{x})$. We denote the resulting directed graph by $G_X(f)$. In this setup, f' refines f simply means $G_X(f')$ is a directed subgraph of $G_X(f)$. We say that a function from X to T is locally constant if it is constant on each connected component of G_X. It is clear that any function from X to T refines a locally constant function. Consequently, the locally constant functions form the greatest element in the partial order $\preceq$. The directed graph of this class, denoted by

$G_X(*)$, is obtained by replacing each edge in G_X by a pair of opposing arcs. On the other hand, the refinement class of a proper coloring of G_X, i.e. a function that assigns distinct elements of T to neighbors in G_X, is minimal in $\preceq$. Let us illustrate these ideas by the following simple example.

Example 3.2 Let X be the Cartesian power $\{0, 1\}^2$ and T be the totally ordered set $\{0, 1\}$ with $0 < 1$. Then G_X is the square graph. The diagram in Figure 2 shows the directed graphs and the relation between the refinement classes of four functions from X to T. The constant function determines the greatest class and the proper coloring

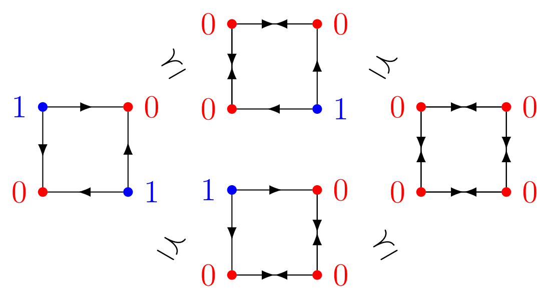


Figure 2: Part of the refinement order

determines a minimal class in the refinement order.

Let $(T, +, 0_T, \leqslant)$ be an ordered abelian group. For $t \in T$, we write $|t|$ for $\max\{t, -t\}$. Let A and B be subsets of T, we write $|A|$ for the set $\{|a| \colon a \in A\}$ and write $A > B$ if $a > b$ for every $a \in A$ and $b \in B$. If f is a function from a Cartesian product X to T, we write $\Delta(f)$ for the set $\{f(\mathbf{x}') - f(\mathbf{x}) \colon \mathbf{x} \text{ and } \mathbf{x}' \text{ are adjacent in } G_X\}$ and call it the **difference set** of f. We write $\Delta^+(f)$ for the set of positive elements of $\Delta(f)$. Note that $\Delta(f) = -\Delta(f)$ and so $\Delta^+(f) = |\Delta(f)| \setminus \{0_T\}$. Consequently, $\Delta^+(f)$ is empty if and only if f is constant on neighbors in G_X if and only if f is locally constant.

Theorem 3.3 *Let $(T, +, 0_T, \leqslant)$ be an ordered abelian group. If f and g are functions from a Cartesian product X to T with $\Delta^+(f) > \Delta^+(g)$, then $f + g$ refines f.*

Proof. Suppose $\mathbf{x}$ and $\mathbf{x}'$ in X differ by one coordinate and that $f(\mathbf{x}') > f(\mathbf{x})$. Since $|g(\mathbf{x}) - g(\mathbf{x}')|$ is either 0_T or in $\Delta^+(g)$, then by assumption it is smaller than $f(\mathbf{x}') - f(\mathbf{x})$. Thus,

$$\begin{aligned} (f+g)(\mathbf{x}') - (f+g)(\mathbf{x}) &= (f(\mathbf{x}') - f(\mathbf{x})) - (g(\mathbf{x}) - g(\mathbf{x}')) \\ &\geqslant (f(\mathbf{x}') - f(\mathbf{x})) - |g(\mathbf{x}) - g(\mathbf{x}')| > 0_T. \end{aligned}$$

This shows that $f + g$ refines f. $\square$

Theorem 3.3 manifests a simple idea: a function can be refined by adding a small (relative to the function) perturbation. The next few propositions should convince the reader that this idea is central to our treatment of refinements.

Proposition 3.4 *Let f, g be real-valued functions on a Cartesian product X. Then $f+\alpha g$ refines f for any $\alpha \in \mathbb{R}$ such that $|\alpha|\Delta^+(g) < \Delta^+(f)$.*

Proof. The proposition is trivial for $\alpha = 0$. For $\alpha \neq 0$,

$$\Delta^+(\alpha g) = |\Delta(\alpha g)| \setminus \{0\} = |\alpha||\Delta(g)| \setminus \{0\} = |\alpha| (|\Delta(g)| \setminus \{0\}) = |\alpha|\Delta^+(g),$$

and so it follows from Theorem 3.3. □

Readers who are familiar with ordered fields will certainly recognize the validity of Proposition 3.4 and Corollary 3.5 below in that setting because the properties of $\mathbb{R}$ being used in their proofs are those of an ordered field.

Corollary 3.5 *Suppose the difference sets of the functions f and g in Proposition 3.4 are finite. Then there exists $\alpha_0 \in (0, \infty]$, depending on both f and g, such that $f + \alpha g$ refines f for any α with $|\alpha| < \alpha_0$.*

Proof. The assumption is equivalent to the sets $\Delta^+(f)$ and $\Delta^+(g)$ being finite. If either of them is empty, then either f or g is constant on neighbors in G_X and so $f + \alpha g$ refines f for any $\alpha \in \mathbb{R}$. We then establish the corollary by taking $\alpha_0 = \infty$. If $\Delta^+(f)$ and $\Delta^+(g)$ are nonempty, then both $m := \min \Delta^+(f)$ and $M := \max \Delta^+(g)$ exist and are positive numbers. So for any α with $|\alpha| < \alpha_0 := m/M$,

$$|\alpha|\Delta^+(g) \leqslant |\alpha|M < m \leqslant \Delta^+(f).$$

Thus, the corollary follows from Proposition 3.4. □

We now turn to the case in which the functions involved are metrics.

Proposition 3.6 *Let d, d_0 be metrics on a set E such that d is $\lambda\mathbb{Z}$-valued and d_0 is $[0, \lambda)$-valued for some positive λ. Then $d + d_0$ is a metric refining d. In particular, if d takes integer values then $d + d'_\flat$ is a metric refining d for any metric d' on E.*

Proof. The sum of two metrics is a metric (Proposition 2.5) so $d + d_0$ is a metric on E. Since

$$\Delta^+(d_0) \subseteq (0, \lambda) < \Delta^+(d) \subseteq \lambda\mathbb{N},$$

it follows from Theorem 3.3 that $d + d_0$ refines d. The special case follows because $d'_\flat$ is a $[0, 1)$-valued metric for any metric d' (Proposition 2.4). □

The metric $d + d_0$ in Proposition 3.6 need not be taking values in a discrete set. So the process of refining metrics given by that proposition cannot be iterated in general. However, Corollary 3.5 can be used instead if the metrics involved have finite difference sets. This happens if the metric space is finite or more generally when the metrics involved take only finitely many values. For instance, given a sequence $d_1, \ldots, d_n$ of metrics on a finite graph G we construct another sequence of metrics $(d'_i)_{1 \leqslant i \leqslant n}$ on G as follows: let $d'_1 = d_1$ and suppose d'_i has been constructed for some $1 \leqslant i \leqslant n$. Then according to Corollary 3.5, we can construct a refinement d'_{i+1} of d'_i by adding to it a suitable positive multiple of d_{i+1}. By Proposition 2.4 and Proposition 2.5, d'_{i+1} is still a metric. Thus, the construction of $(d'_i)_{1\leqslant i\leqslant n}$ is completed by induction. All metrics in the original sequence contribute to d'_n the last metric in the new sequence. Moreover, the new metrics respect the sequence's order in the sense that for any $1 \leqslant i \leqslant n-1$ the relative closeness of vertices in G determined by d'_i will not be contradicted by d'_{i+1}.

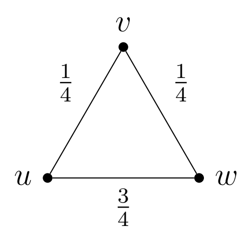
4 Refinements of the Shortest-Path Distance

We now focus on refining a particular metric—the shortest-path distance. General results about refinements in the previous section still apply. The difference here is that the refinements are obtained by modifying the shortest-path distance by functions that are not metrics themselves and the challenge is to come up with the right kind of functions so that the resulting refinements are still metrics.

Let G be a finite simple connected graph. Let w be a non-negative real function on the edges of G. The **product weight** of a path γ in G, denoted by $p(\gamma)$, is defined to be the product of the weights of its edges. That is, $p(\gamma) = \prod_{e\in\gamma} w(e)$. For vertices u and v of G, let

$$\pi(u,v) = \begin{cases} \min\left\{\dfrac{p(\gamma)}{p(\gamma)+1} : \gamma \text{ is a shortest path between } u \text{ and } v.\right\} & \text{if } u \neq v; \\ 0 & \text{if } u = v. \end{cases}$$

Clearly $\pi(u,v)$ is symmetric and takes values in $[0,1)$ but π is not a metric since $\pi(u,v)$ could be 0 for distinct u and v. Moreover, π does not satisfy the triangle inequality in general. For instance, consider the following triangle with the indicated edge weights:



The triangle inequality is violated because $\pi(u,w) = 3/7 > 2/5 = \pi(u,v) + \pi(v,w)$. In the following, let d_s denote the shortest-path distance and let d^π_w be $d_s + \pi$.

Theorem 4.1 *d^π_w is a metric on G.*

Proof. Only the triangle inequality deserves a proof. If it fails for d^π_w, then there exist vertices x, y, z of G such that $d^\pi_w(x,y) > d^\pi_w(x,z) + d^\pi_w(z,y)$. That is,

$$d_s(x,y) + \pi(x,y) > d_s(x,z) + \pi(x,z) + d_s(z,y) + \pi(z,y). \tag{1}$$

Since d_s itself satisfies the triangle inequality and π takes values in $[0,1)$, we obtain the following inequalities by rearranging the terms in (1):

$$1 \geqslant \pi(x,y) - \pi(x,z) - \pi(z,y) > d_s(x,z) + d_s(z,y) - d_s(x,y) \geqslant 0. \tag{2}$$

Because d_s takes integer values, it follows from (2) that

$$d_s(x,y) = d_s(x,z) + d_s(z,y). \tag{3}$$

Choose a shortest path γ_{xz} between x and z such that $p(\gamma_{xz})/(p(\gamma_{xz})+1)$ realizes the value $\pi(x,z)$. Choose γ_{zy} analogously. Equation (3) implies the walk γ obtained by concatenating γ_{xz} and γ_{zy} has length $d_s(x,y)$ and because every walk contains a path with the same ends, γ must be a shortest path between x and y. From this and the fact that p is a nonnegative function, we conclude that

$$\begin{aligned} \pi(x,y) &\leqslant \frac{p(\gamma)}{p(\gamma)+1} = \frac{p(\gamma_{xz})p(\gamma_{zy})}{p(\gamma_{xz})p(\gamma_{zy})+1} \\ &\leqslant \frac{p(\gamma_{xz})}{p(\gamma_{xz})+1} + \frac{p(\gamma_{xz})}{p(\gamma_{xz})+1} = \pi(x,z) + \pi(z,y). \end{aligned} \tag{4}$$

However, the inequalities in (4) are in contradiction with those in (2). Thus, d_w^π must satisfy the triangle inequality as well. □

Since $\Delta^+(\pi) \subseteq (0,1) < \Delta^+(d_s) \subseteq \mathbb{N}$, d_w^π refines d_s by Theorem 3.3. This fact together with Theorem 4.1 imply:

Proposition 4.2 *d_w^π is a metric refining d_s.*

Theorem 4.1 can be generalized in a number of ways. First, the finiteness assumption on G, which is harmless to our applications because the collaboration graph itself is finite, can be dropped. Its only use is to guarantee the minimum in the definition of π exists. For that matter we can assume between any two vertices of G there are only finitely many paths, or even just finitely many shortest paths. In fact, by taking infimum instead of minimum in the definition of π we can drop these assumptions altogether. However, with that change we can no longer guarantee a value of π is realized by a path, yet for any distinct vertices u, v and $\varepsilon > 0$, there will be a shortest path γ_{uv} between u and v with $p(\gamma_{uv})/(p(\gamma_{uv})+1) < \pi(u,v) + \varepsilon$. Since π still takes values in $[0,1)$, the inequalities in (2) continue to hold. Therefore, the walk $\gamma = \gamma_{xz}\gamma_{zy}$ must again be a shortest path between x and y, and so

$$\pi(x,y) \leqslant \frac{p(\gamma)}{p(\gamma)+1} \leqslant \frac{p(\gamma_{xz})}{p(\gamma_{xz})+1} + \frac{p(\gamma_{zy})}{p(\gamma_{zy})+1} < \pi(x,z) + \pi(z,y) + 2\varepsilon.$$

This establishes $\pi(x,y) \leqslant \pi(x,z) + \pi(z,y)$ and that is what is needed to complete the proof.

Second, the assumption on G being connected can also be removed. We extend d_s and π by declaring $d_s(u,v) = \infty$ and $\pi(u,v) = 0$ whenever u, v are in different components of G. With that change d_s is no longer real-valued and hence not a metric in the strict sense. However, if one adopts the usual conventions: $\infty + \infty = \infty$, $r < \infty$ and $\infty + r = \infty$ ($r \in \mathbb{R}$), then d_s still satisfies the triangle inequality. Moreover, $d_w^\pi(u,v) = d_s(u,v) + \pi(u,v) = \infty$ if and only if u and v are in different components of G. It follows that d_w^π can only fail the triangle inequality because it fails on some component of G, i.e. there exist vertices x, y, z in the same component of G with $d_w^\pi(x,y) > d_w^\pi(x,z) + d_w^\pi(z,y)$. Hence, the proof of Theorem 4.1 goes through without any modification.

5 Some Refined Erdős numbers

In this section we propose two different refinements of the shortest-path distance of the collaboration graph[3] C. To demonstrate how these refinements differentiate people who have the same Erdős number, we compute the new Erdős numbers of a few mathematicians. For an edge e in C, let its weight $w(e)$ be the reciprocal of $j(e)$ the number of joint articles between the two ends of e. As in Section 4, d_s denotes the shortest-path distance and d_w^π denotes the refinement $d_s + \pi$. The **sum weight** of a path γ, denoted by $s(\gamma)$, is defined to be the sum of the weights of its edges. That is, $s(\gamma) = \sum_{e\in\gamma} w(e)$. The **lightest-path distance** on C is then given by the function

$$\sigma(u,v) = \begin{cases} \min\{s(\gamma) \colon \gamma \text{ is a path between } u \text{ and } v.\} & \text{if } u \neq v; \\ 0 & \text{if } u = v. \end{cases} \tag{5}$$

We write d_w^σ for the refinement $d_s + \sigma_\flat$ of d_s and use $\mathrm{pEN}(x)$ and $\mathrm{sEN}(x)$ to denote the Erdős number of x defined by d_w^π and d_w^σ, respectively[4].

We rely on the data provided by MathSciNet and the Erdős Number Project for computations. By definition, Erdős himself has pEN and sEN zero. András Sárközy is the most frequent collaborator of Erdős and vice versa. They have 62 joint papers. It follows that Sárközy has the smallest positive refined Erdős numbers of all. The pEN and the sEN of Sárközy are both

$$1 + \frac{\frac{1}{62}}{1+\frac{1}{62}} = \frac{64}{63} \approx 1.016.$$

The most frequent collaborator of András Hajnal is Erdős. They have 57 joint publications. Consequently, both pEN and sEN of Hajnal are $58/57 \approx 1.017$. Hajnal is the second most frequent collaborator of Erdős. He also has the second smallest positive pEN and sEN.

Christian Mauduit has Erdős number 1 and has two joint articles with Erdős. So his pEN is $4/3$. Mauduit's most frequent collaborator is Sárközy, they have 41 joint papers. His second most frequent collaborator is Joël Rivat. They co-authored 16 articles. The path Mauduit–Sárközy–Erdős has sum weight $1/41 + 1/62 = 103/2542$ which is less than the weight of the edge $(1/16)$ between Mauduit and Rivat. Therefore, it must be the lightest path between Mauduit and Erdős. Thus,

$$\mathrm{sEN}(\mathrm{Mauduit}) = 1 + \frac{\frac{103}{2542}}{1+\frac{103}{2542}} = \frac{2748}{2645} \approx 1.039.$$

István Juhász has Erdős number 2. The six shortest paths between Juhász and Erdős go through András Hajnal, Peter Hamburger, Kenneth Kunen, Menachem Magidor, Mary Ellen Estill Rudin and Saharon Shelah, respectively. We organize the information given by these paths into Table 1.

[3]It is the graph given by the MathSciNet database as of the time of submission of this article.

[4]EN stands for Erdős Number. The 'p' and the 's' in the notation indicate the refinements are coming from a product and a sum of edge weights, respectively.

$j(\text{Juhász}, x)$	32	1	5	1	1	13
x	Hajnal	Hamburger	Kunen	Magidor	Rudin	Shelah
$j(x, \text{Erdős})$	57	1	1	1	1	3

Table 1: The shortest paths between Juhász and Erdős

From the table it is clear that

$$\text{pEN(Juhász)} = 2 + \frac{\frac{1}{57}\frac{1}{32}}{1 + \frac{1}{57}\frac{1}{32}} = \frac{3651}{1825} \approx 2.001.$$

Given the form of data available to us it is harder to compute the sEN of Juhász as it is more difficult to determine which paths between Juhász and Erdős are the lightest. First, the path Juhäsz–Hajnal–Erdős gives $\frac{1}{57} + \frac{1}{32} = \frac{89}{1824} \approx 0.049$ as an upper bound of the lightest-path distance between Juhász and Erdős. Since $1824/89 \approx 20.5$, we only need to examine the collaborators of Juhász, besides Hajnal, who have at least 21 joint articles with Juhász. Only two mathematicians, namely Lajos Soukup (with 29 joint papers) and Zoltán Szentmiklóssy (with 44 joint papers) meet this requirement. Solving the inequality

$$\frac{89}{1824} > \frac{1}{44} + \frac{1}{n}$$

yields $n > 38$. So in order for a lightest path between Juhász and Erdős to go through either Szentmiklóssy or Soukup, each of them needs a collaborator other than Juhász with at least 39 joint articles. It turns out that the most frequent collaborator of both Soukup and Szentmiklóssy is Juhász and their second most frequent collaborators are each other. They have 24 joint articles. From this we conclude that Juhász–Hajnal–Erdős is the unique lightest path between Juhász and Erdős. Consequently,

$$\text{sEN(Juhász)} = 2 + \frac{\frac{89}{1824}}{1 + \frac{89}{1824}} = \frac{3915}{1913} \approx 2.047.$$

We summarize this information[5] in Figure 3. Moreover, Soukup and Szentmiklóssy both have 10 joint articles with their third most frequent collaborators (János Gerlits for Szentmiklóssy and Saharon Shelah for Soukup). With this additional information, we can compute their sEN's and pEN's. We skip the details here but summarize in Table 2 the new Erdős numbers of the mathematicians appearing in Figure 3.

We conclude this article by suggesting another edge weight function for the collaboration graph that seems appropriate for the purpose of measuring closeness between authors: take the weight of an edge to be the ratio $j(e)/t(e)$ where $j(e)$ is the number of joint articles and $t(e)$ is the total number of articles published by the ends of e. A more sophisticated version of this weighting function that takes the types of publication into account has been considered recently in [10]. We leave the computations of the corresponding refined Erdős numbers to the interested reader.

[5]Here an edge is labeled not by its weight but by the number of joint articles of its ends.

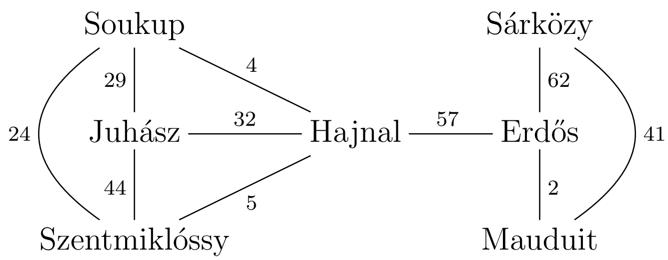


Figure 3: A subgraph of the collaboration graph C.

x	EN	pEN(x)	sEN(x)
Erdős	0	0	0
Sárközy	1	$64/63 \approx 1.016$	$64/63 \approx 1.016$
Hajnal	1	$59/58 \approx 1.017$	$59/58 \approx 1.017$
Mauduit	1	$4/3 \approx 1.333$	$2748/2645 \approx 1.039$
Juhász	2	$3651/1825 \approx 2.001$	$3915/1913 \approx 2.047$
Szentimiklóssy	2	$573/286 \approx 2.003$	$44433/21499 \approx 2.067$
Soukup	2	$459/229 \approx 2.004$	$119007/57301 \approx 2.077$

Table 2: Refined Erdős Numbers of a few mathematicians.

Acknowledgments

We thank Serban Raianu for carefully reading an earlier version of this article and for bringing [8] to our attention. We also thank him and the referees for valuable advice which helped improve the presentation.

References

[1] A.T. Balaban, D.J. Klein, Co-authorship, rational Erdős numbers, and resistance distances in graphs, *Scientometrics*, **55** (2002), 59–70.

[2] M. Barr, Rational Erdős Number, available online at the URL:`https://www.oakland.edu/Assets/upload/docs/Erdos-Number-Project/barr.pdf`.

[3] F.L. Bauer, J. Stoer, C. Witzgall, Absolute and monotonic norms, *Numer. Math.*, **3** (1961), 257–264.

[4] B. Bollobás, *Modern graph theory*, volume 184 of Graduate Texts in Mathematics, Springer-Verlag, 1998.

[5] P. Chebotarev, A class of graph-geodetic distances generalizing the shortest-path and the resistance distances, *Discrete Appl. Math.*, **159** (2011), 295–302.

[6] S. Dasgupta, C.H. Papadimitriou, U.V. Vazirani, *Algorithms*, McGraw-Hill, 2008.

[7] D. Gries, Characterizations of certain classes of norms. *Numer. Math.*, **10** (1967), 30–41.

[8] G. Morrison, L. Dudte, L. Mahadevan, Generalized Erdős numbers for network analysis, *R. Soc. Open Sci.*, **5** (2018), 22pp, available online at the URL:`https://royalsocietypublishing.org/doi/10.1098/rsos.172281`.

[9] L.W. Shapiro, An electrical lemma, *Math. Mag.*, **60** (1987), 36–38.

[10] V. Ströele, R. Crivano, G. Zimbrão, J.M. Souze, F. Campos, J.M.N. David, R. Braga, Rational Erdős number and maximum flow as measurement models for scientific social network analysis, *J. Braz. Comput. Soc.*, **24** (2018), 17pp, available online at the URL:`https://link.springer.com/article/10.1186/s13173-018-0070-6`.

[11] Y. Yang, D.J. Klein, A recursion formula for resistance distances and its applications, *Discrete Appl. Math.*, **161** (2013), 2702–2715, available online at the URL:`https://doi.org/10.1016/j.dam.2012.07.015`.

Kayla Lock
Arizona State University
1151 S Forest Ave, Tempe, AZ 85281
E-mail: `kalock@asu.edu`

Wai Yan Pong
California State University, Dominguez Hills
1000 E Victoria Street, Carson, CA 90747
E-mail: `wpong@csudh.edu`

Alexander Wittmond
University of Missouri
Columbia, MO 65211
E-mail: `ajwcz7@mail.missouri.edu`

Received: May 23, 2019 **Accepted**: August 13, 2019
Communicated by Serban Raianu